

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:41:26 ; Search time 33.8015 Seconds
(without alignments)
42.263 Million cell updates/sec

Title: US-10-089-549-1

Perfect score: 43

Sequence: 1 YVSLGNLI 9

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: A Geneseq 19Jun03.*

```

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	100.0	9	23 ABB84268	B. globisporus deh
2	43	100.0	9	23 ABG30540	Alpha-isomaltosylg
3	43	100.0	9	23 AAO17330	Alpha-isomaltosylg
4	43	100.0	1249	23 ABG30537	Alpha-isomaltosylg
5	43	100.0	1284	23 ABG30563	Alpha-isomaltosylg
6	43	100.0	1284	24 ABP70652	Bacillus globispor
7	43	100.0	1284	24 ABP57724	Protein #1 related
8	35	81.4	154	16 AAR80184	Mispi-derived poly
9	35	81.4	773	22 AAB48343	S. pneumoniae Sp13

10	35	81.4	831	16 AAR80168	pMIS11 MiSP spider
11	35	81.4	2120	21 AAY81710	Streptococcus pneu
12	35	81.4	2140	24 ABU01020	S. pneumoniae type
13	33	76.7	9	23 AAO17340	Alpha-isomaltosylg
14	33	76.7	1251	23 ABG30538	Alpha-isomaltosylg
15	33	76.7	1286	23 ABG30564	Alpha-isomaltosylg
16	32	74.4	188	22 ABG48346	Human liver peptid
17	32	74.4	188	22 ABB28324	Human liver peptid
18	32	74.4	188	22 ABB33500	Peptide #1006 enco
19	32	74.4	188	22 ABB18960	Protein #959 enco
20	32	74.4	188	22 AAM54280	Human brain expres
21	32	74.4	188	22 AAM66675	Human bone marrow
22	32	74.4	188	22 AAM14543	Peptide #377 enco
23	32	74.4	188	22 AAM28960	Peptide #997 enco
24	32	74.4	188	22 AAM02269	Peptide #951 enco
25	32	74.4	188	23 ABG36329	Human peptide enco
26	32	74.4	195	23 ABG59989	Human DITHP polype
27	32	74.4	201	24 ABE33686	Human structural a
28	32	74.4	282	21 AAG44489	Arabidopsis thalia
29	32	74.4	283	22 AAB30843	Amino acid sequenc
30	32	74.4	298	21 AAG44488	Arabidopsis thalia
31	32	74.4	298	22 AAB30835	Amino acid sequenc
32	32	74.4	312	22 ABB63330	Drosophila melanog
33	32	74.4	312	22 AAB60378	Drosophila atonal
34	32	74.4	332	22 AAG66992	Formanidopyrimidin
35	32	74.4	346	22 AAB73172	Human HA-LPA-R #1.
36	32	74.4	356	20 AAY38890	Neisseria meningit
37	32	74.4	356	20 AAY38891	Neisseria meningit
38	32	74.4	356	20 AAY38893	Neisseria gonorrhoe
39	32	74.4	356	24 ABP79218	N. gonorrhoeae ami
40	32	74.4	369	21 AAY96486	G-protein coupled
41	32	74.4	369	22 AAE04551	Human G-protein co
42	32	74.4	374	23 ABP71953	Candida albicans e
43	32	74.4	393	22 AAB73173	Human HA-LPA-R #2.
44	32	74.4	415	22 AAB30842	Amino acid sequenc
45	32	74.4	419	20 AAY14195	Human lysophosphati

ALIGNMENTS

RESULT 1

ID ABB84268 standard; peptide; 9 AA.

XX ABB84268;

XX 08-OCT-2002 (first entry)

DT B. globisporus dehydration-associated peptide SEQ ID 1.

DE Dehydration; cyclo-alpha-D-glucopyranosyl; pharmaceutical; agrochemical;

KW toletry; non-reducing saccharide.

OS Bacillus globisporus.

PN WO200257011-A1.

XX 25-JUL-2002.

PF 17-JAN-2002; 2002WO-JP00288.

XX 19-JAN-2001; 2001JP-0010991.

XX (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.

PI Kubota M, Nishimoto T, Aga H, Fukuda S, Miyake T;

XX WPI; 2002-590706/63.

PT Dehydrating agent used for pharmaceuticals, agrochemicals and

XX toiletries comprises cyclic tetra-saccharide -

PS Disclosure; Page 134; 140pp; Japanese.

XX This invention describes a novel dehydrating agent comprising
CC cyclo(6)-alpha-D-glucopyranosyl-(1-3)-alpha-D-glucopyranosyl-(1-6)-alpha-
CC D-glucopyranosyl-(1-3)-alpha-Glucopyranosyl. The agent of the invention
CC can be used in pharmaceuticals, agrochemicals and toiletries. The agent
CC is a non-reducing saccharide and dries articles without causing a
CC deterioration in quality of the articles. This sequence represents a
CC peptide described in the disclosure of the invention.
XX

XX Sequence 9 AA;

Query Match 100.0%; Score 43; DB 23; Length 9;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YVSSLGNI 9

DB 1 YVSSLGNI 9

RESULT 2

ABG30540

ID ABG30540 standard; Peptide; 9 AA.

AC ABG30540;

XX

XX 07-OCT-2002 (first entry)

XX

DE Alpha-isomaltosylglucosaccharide synthase related peptide #1.

XX

XX Alpha-isomaltosylglucosaccharide synthase; sugar production;

KW cyclic tetrasaccharide; cyclo(6)-alpha-D-glucopyranosyl(1-3);

KW alpha-isomaltosyltransferase; viscosity modifier; sweetener;

KW heat stabiliser; hydrolysis stabiliser; colour stabiliser; foodstuff;

KW cosmetic; drug composition.

XX Unidentified.

OS

XX WC200255708-A1.

PN

XX 18-JUL-2002.

XX

XX 09-JAN-2002; 2002WO-JP00052.

XX

XX 12-JAN-2001; 2001JP-0005441.

XX

XX (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.

PA

XX Kubota M, Matura K, Yamamoto T, Fukuda S;

PI

XX WPI; 2002-520129/55.

XX

XX New alpha-isomaltosylglucosaccharide synthase of bacterial origin for

PT the production of cyclic tetrasaccharide gum

PT

XX Disclosure; Page 117; 144pp; Japanese.

XX

XX The invention describes novel microbial polypeptides having

CC alpha-isomaltosylglucosaccharide synthase activity. The proteins are

CC useful for producing a sugar (I) having at least three glucose units

CC connected by alpha-1,4 bonds except for alpha-1,6 at the non-reducing

CC end. The invention also describes a method for the production of the

CC cyclic tetrasaccharide cyclo(6)-alpha-D-glucopyranosyl(1-3)-alpha-D-

CC glucopyranosyl(1-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-glucopyranosyl

CC (1-1) (II) by treatment of (I) with alpha-isomaltosyltransferase.

CC (II) and similar sugars in crystalline or syrup form are used as

CC a viscosity modifier, sweetener, heat stabiliser, hydrolysis stabiliser

CC and colour stabiliser for foodstuffs, cosmetics and drug compositions.

CC This is the amino acid sequence of an alpha-isomaltosylglucosaccharide

XX synthase related peptide.

XX Sequence 9 AA;

Query Match

Best Local Similarity 100.0%; Score 43; DB 23; Length 9;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YVSSLGNI 9

DB 1 YVSSLGNI 9

RESULT 3

AAO17330

ID AAO17330 standard; peptide; 9 AA.

XX

XX AAO17330;

XX

XX 08-JUL-2002 (first entry)

XX

XX Alpha-isomaltosylglucosaccharide synthase related peptide #1.

XX

KW Alpha-isomaltosylglucosaccharide synthase; enzyme; sweetener; food;

KW drink; cosmetics; pharmaceutical; cyclic tetrasaccharide.

XX

OS Bacillus globisporus.

XX

XX WC200210361-A1.

PN

XX 07-FEB-2002.

XX

XX 25-JUL-2001; 2001WO-JP06412.

XX

XX 01-AUG-2000; 2000JP-0233364.

XX

XX 02-AUG-2000; 2000JP-0234937.

XX

XX (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.

PA

XX Kubota M, Teusaki K, Higashiyama T, Fukuda S, Miyake T;

PI

XX WPI; 2002-315302/35.

XX

XX Alpha-Isomaltosylglucosaccharide synthase capable of transferring

PT Alpha-glucosyl from saccharide, useful in producing cyclic

PT tetrasaccharides and related carbohydrates industrially for application

PT in compositions e.g. as drugs

XX

XX Claim 5; Page 200; 209pp; Japanese.

XX

XX The present invention relates to an alpha-isomaltosylglucosaccharide

CC synthase which is capable of transferring alpha-glucosyl from a

CC saccharide to form a specific saccharide which carries an alpha-1,6

CC glucosyl linkage at its non-reducing end and an alpha-1,4 glucosyl

CC linkage at ends other than the non-reducing end and has a degree of

CC glucose polymerization of at most 3, but without substantially elevating

CC the reducing ability. The synthase is useful in producing cyclic

CC tetrasaccharides and related carbohydrates industrially for application

CC in compositions e.g. as sweeteners, low calorie food materials, taste

CC improvers, flavour improvers, quality improvers, water-separation

CC inhibitors, stabilizers, excipients, binding agents and pulverisation

CC bases, especially in foods, drinks, cosmetics and pharmaceuticals. The

CC present sequence is a peptide described in the exemplification of the

XX invention.

XX Sequence 9 AA;

QY

1 YVSSLGNI 9

DB

1 YVSSLGNI 9

Query Match 100.0%; Score 43; DB 23; Length 9;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 4
ABG30537
ID ABG30537 standard; Protein; 1249 AA.
XX
AC ABG30537;
XX
DT 07-OCT-2002 (first entry)
XX
DE Alpha-isomaltosylglucosaccharide synthase #1 mature protein.
XX
KW Alpha-isomaltosylglucosaccharide synthase; sugar production;
KW cyclic tetrasaccharide; cyclo(-6)-alpha-D-glucopyranosyl(1-3);
KW alpha-isomaltosyltransferase; viscosity modifier; sweetener;
KW heat stabiliser; hydrolysis stabiliser; colour stabiliser; foodstuff;
KW cosmetic; drug composition; enzyme.
XX
OS Unidentified.
XX
PN WO200255708-A1.
XX
PD 18-JUL-2002.
XX
PF 09-JAN-2002; 2002WO-JP00052.
XX
PR 12-JAN-2001; 2001JP-0005441.
XX
PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.
XX
PI Kubota M, Maruta K, Yamamoto T, Fukuda S;
XX
DR WPI; 2002-520129/55.
XX
DR N-PSDB; ABK88154.
XX
PT New alpha-isomaltosylglucosaccharide synthase of bacterial origin for
the production of cyclic tetrasaccharide gum
XX
PS Claim 1; Page 101-105; 144pp; Japanese.
XX
CC The invention describes novel microbial polypeptides having
alpha-isomaltosylglucosaccharide synthase activity. The proteins are
useful for producing a sugar (I) having at least three glucose units
connected by alpha-1,4 bonds except for alpha-1,6 at the non-reducing
end. The invention also describes a method for the production of the
cyclic tetrasaccharide cyclo(-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-
glucopyranosyl(1-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-glucopyranosyl
(1-3) (II) by treatment of (I) with alpha-isomaltosyltransferase.
CC (I) and similar sugars in crystalline or syrup form are used as
a viscosity modifier, sweetener, heat stabiliser, hydrolysis stabiliser
and colour stabiliser for foodstuffs, cosmetics and drug compositions.
CC This is the amino acid sequence of an alpha-isomaltosylglucosaccharide
synthase mature protein.
XX
SQ Sequence 1249 AA;
Query Match 100.0%; Score 43; DB 23; Length 1249;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 YVSSIGNLI 9
DB 1 YVSSIGNLI 9
RESULT 5
ABG30563
ID ABG30563 standard; Protein; 1284 AA.
XX
AC ABG30563;
XX
DT 07-OCT-2002 (first entry)
XX
DE Alpha-isomaltosylglucosaccharide synthase #1.
XX

XX
KW Alpha-isomaltosylglucosaccharide synthase; sugar production;
KW cyclic tetrasaccharide; cyclo(-6)-alpha-D-glucopyranosyl(1-3);
KW alpha-isomaltosyltransferase; viscosity modifier; sweetener;
KW heat stabiliser; hydrolysis stabiliser; colour stabiliser; foodstuff;
KW cosmetic; drug composition.
XX
OS Unidentified.
XX
PN WO200255708-A1.
XX
PD 18-JUL-2002.
XX
PF 09-JAN-2002; 2002WO-JP00052.
XX
PR 12-JAN-2001; 2001JP-0005441.
XX
PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.
XX
PI Kubota M, Maruta K, Yamamoto T, Fukuda S;
XX
DR WPI; 2002-520129/55.
XX
DR N-PSDB; ABK88157.
XX
PT New alpha-isomaltosylglucosaccharide synthase of bacterial origin for
the production of cyclic tetrasaccharide gum
XX
PS Disclosure; Page 120-136; 144pp; Japanese.
XX
CC The invention describes novel microbial polypeptides having
alpha-isomaltosylglucosaccharide synthase activity. The proteins are
useful for producing a sugar (I) having at least three glucose units
connected by alpha-1,4 bonds except for alpha-1,6 at the non-reducing
end. The invention also describes a method for the production of the
cyclic tetrasaccharide cyclo(-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-
glucopyranosyl(1-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-glucopyranosyl
(1-3) (II) by treatment of (I) with alpha-isomaltosyltransferase.
CC (I) and similar sugars in crystalline or syrup form are used as
a viscosity modifier, sweetener, heat stabiliser, hydrolysis stabiliser
and colour stabiliser for foodstuffs, cosmetics and drug compositions.
CC This is the amino acid sequence of an alpha-isomaltosylglucosaccharide
synthase.
XX
SQ Sequence 1284 AA;
Query Match 100.0%; Score 43; DB 23; Length 1284;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 YVSSIGNLI 9
DB 36 YVSSIGNLI 44
RESULT 6
ABP70652
ID ABP70652 standard; Protein; 1284 AA.
XX
AC ABP70652;
XX
DT 06-MAR-2003 (first entry)
XX
DE Bacillus globisporus protein encoded by SEQ ID 2.
XX
KW Bacillus globisporus.
XX
OS Cyclic tetrasaccharide; sweetening agent.
XX
FH Key Location/Qualifiers
FT Peptide 1..35
FT /label= signal_peptide
XX
PN WO200272594-A1.
XX

CC The invention relates to a novel method for the production of isomaltose,
CC comprising treating a saccharide having alpha-1,4-glucosyl linkage at the
CC non-reducing end and a glucose polymerization degree of up to 2 with
CC alpha-isomaltosyl glucosaccharide synthase, treating the product(s) with
CC isomaltose-releasing enzyme and collecting the obtained isomaltose.
CC Isomaltose is useful for manufacturing beverages, health foods, feeding
CC materials, cosmetics, drugs and cooking sauces. The sequence represents a
CC polypeptide used in the invention. No further information about the
CC sequence can be obtained from the specification.
XX SQ Sequence 1284 AA;

Query Match 100.0%; Score 43; DB 24; Length 1284;
Best Local Similarity 100.0%; Pred. No. 6.5; Mismatches 0; Indels 0; Gaps 0;
Matches 9; Conservative 0;

QY 1 YVSSLGNI 9
DB 36 YVSSLGNI 44
RESULT 8
AAR80184
ID AAR80184 standard; peptide; 154 AA.
XX AC AAR80184;
XX AC AAR80184;
XX DT 30-APR-1996 (first entry)
XX DE MISPI-derived polypeptide amino terminal.
XX KM Spider silk; repeat unit; consensus; minor ampullate silk protein;
XX KM spidroin; MISPI; orb web spider; dragline.
XX OS Nepheila clavipes.
XX PN WO9525165-A1.
XX PD 21-SEP-1995.
XX PF 14-MAR-1995; 95WO-US03139.
XX PR 14-MAR-1994; 94US-0209747.
XX PA (UWVY-) UNIV WYOMING.
XX PI Colgin M, Lewis RV;
XX DR WPI; 1995-336970/43.
XX PT Polypeptide(s) comprising repeated unit amino acid sequences, also
XX PT cDNAs - derived from minor ampullate spider silk proteins and used
XX PT to form spider silk fibres
XX PS Claim 5; Page 60; 86pp; English.
XX CC A cDNA clone (pMIS1) encoding the orb web spider minor ampullate silk
XX CC protein MISPI has been identified and sequenced. MISPI is made up of
XX CC polypeptide repeat units (see R8016983). The repeat unit peptides may
XX CC form part of a larger polypeptide contg. an amino terminus
XX CC (see AAR80184-85) and a carboxy terminus (see AAR80186-89). Repeat unit
XX CC peptides make up spider silk proteins (spidroins) which in turn
XX CC aggregate to form the silk fibres. Spider silk fibres have high tensile
XX CC strength and significant elasticity. An isolated cDNA clone of a silk
XX CC protein encoding sequence is of use to produce the protein at high
XX CC yields using recombinant DNA technology.
XX SQ Sequence 154 AA;

Query Match 81.4%; Score 35; DB 16; Length 154;
Best Local Similarity 77.8%; Pred. No. 29; Mismatches 1; Indels 0; Gaps 0;
Matches 7; Conservative 1;

QY 1 YVSSLGNI 9
DB 36 YVSSLGNI 44
RESULT 8
AAR80184
ID AAR80184 standard; peptide; 154 AA.
XX AC AAR80184;
XX AC AAR80184;
XX DT 30-APR-1996 (first entry)
XX DE MISPI-derived polypeptide amino terminal.
XX KM Spider silk; repeat unit; consensus; minor ampullate silk protein;
XX KM spidroin; MISPI; orb web spider; dragline.
XX OS Nepheila clavipes.
XX PN WO9525165-A1.
XX PD 21-SEP-1995.
XX PF 14-MAR-1995; 95WO-US03139.
XX PR 14-MAR-1994; 94US-0209747.
XX PA (UWVY-) UNIV WYOMING.
XX PI Colgin M, Lewis RV;
XX DR WPI; 1995-336970/43.
XX PT Polypeptide(s) comprising repeated unit amino acid sequences, also
XX PT cDNAs - derived from minor ampullate spider silk proteins and used
XX PT to form spider silk fibres
XX PS Claim 5; Page 60; 86pp; English.
XX CC A cDNA clone (pMIS1) encoding the orb web spider minor ampullate silk
XX CC protein MISPI has been identified and sequenced. MISPI is made up of
XX CC polypeptide repeat units (see R8016983). The repeat unit peptides may
XX CC form part of a larger polypeptide contg. an amino terminus
XX CC (see AAR80184-85) and a carboxy terminus (see AAR80186-89). Repeat unit
XX CC peptides make up spider silk proteins (spidroins) which in turn
XX CC aggregate to form the silk fibres. Spider silk fibres have high tensile
XX CC strength and significant elasticity. An isolated cDNA clone of a silk
XX CC protein encoding sequence is of use to produce the protein at high
XX CC yields using recombinant DNA technology.
XX SQ Sequence 154 AA;

Query Match 81.4%; Score 35; DB 16; Length 154;
Best Local Similarity 77.8%; Pred. No. 29; Mismatches 1; Indels 0; Gaps 0;
Matches 7; Conservative 1;

QY 1 YVSSLGNI 9
DB 36 YVSSLGNI 44
RESULT 8
AAR80184
ID AAR80184 standard; peptide; 154 AA.
XX AC AAR80184;
XX AC AAR80184;
XX DT 30-APR-1996 (first entry)
XX DE MISPI-derived polypeptide amino terminal.
XX KM Spider silk; repeat unit; consensus; minor ampullate silk protein;
XX KM spidroin; MISPI; orb web spider; dragline.
XX OS Nepheila clavipes.
XX PN WO9525165-A1.
XX PD 21-SEP-1995.
XX PF 14-MAR-1995; 95WO-US03139.
XX PR 14-MAR-1994; 94US-0209747.
XX PA (UWVY-) UNIV WYOMING.
XX PI Colgin M, Lewis RV;
XX DR WPI; 1995-336970/43.
XX PT Polypeptide(s) comprising repeated unit amino acid sequences, also
XX PT cDNAs - derived from minor ampullate spider silk proteins and used
XX PT to form spider silk fibres
XX PS Claim 5; Page 60; 86pp; English.
XX CC A cDNA clone (pMIS1) encoding the orb web spider minor ampullate silk
XX CC protein MISPI has been identified and sequenced. MISPI is made up of
XX CC polypeptide repeat units (see R8016983). The repeat unit peptides may
XX CC form part of a larger polypeptide contg. an amino terminus
XX CC (see AAR80184-85) and a carboxy terminus (see AAR80186-89). Repeat unit
XX CC peptides make up spider silk proteins (spidroins) which in turn
XX CC aggregate to form the silk fibres. Spider silk fibres have high tensile
XX CC strength and significant elasticity. An isolated cDNA clone of a silk
XX CC protein encoding sequence is of use to produce the protein at high
XX CC yields using recombinant DNA technology.
XX SQ Sequence 154 AA;

Query Match 81.4%; Score 35; DB 16; Length 154;
Best Local Similarity 77.8%; Pred. No. 29; Mismatches 1; Indels 0; Gaps 0;
Matches 7; Conservative 1;

QY 1 YVSSLGNI 9
DB 36 YVSSLGNI 44
RESULT 8
AAR80184
ID AAR80184 standard; peptide; 154 AA.
XX AC AAR80184;
XX AC AAR80184;
XX DT 30-APR-1996 (first entry)
XX DE MISPI-derived polypeptide amino terminal.
XX KM Spider silk; repeat unit; consensus; minor ampullate silk protein;
XX KM spidroin; MISPI; orb web spider; dragline.
XX OS Nepheila clavipes.
XX PN WO9525165-A1.
XX PD 21-SEP-1995.
XX PF 14-MAR-1995; 95WO-US03139.
XX PR 14-MAR-1994; 94US-0209747.
XX PA (UWVY-) UNIV WYOMING.
XX PI Colgin M, Lewis RV;
XX DR WPI; 1995-336970/43.
XX PT Polypeptide(s) comprising repeated unit amino acid sequences, also
XX PT cDNAs - derived from minor ampullate spider silk proteins and used
XX PT to form spider silk fibres
XX PS Claim 5; Page 60; 86pp; English.
XX CC A cDNA clone (pMIS1) encoding the orb web spider minor ampullate silk
XX CC protein MISPI has been identified and sequenced. MISPI is made up of
XX CC polypeptide repeat units (see R8016983). The repeat unit peptides may
XX CC form part of a larger polypeptide contg. an amino terminus
XX CC (see AAR80184-85) and a carboxy terminus (see AAR80186-89). Repeat unit
XX CC peptides make up spider silk proteins (spidroins) which in turn
XX CC aggregate to form the silk fibres. Spider silk fibres have high tensile
XX CC strength and significant elasticity. An isolated cDNA clone of a silk
XX CC protein encoding sequence is of use to produce the protein at high
XX CC yields using recombinant DNA technology.
XX SQ Sequence 154 AA;

2D 19-SEP-2002.
XX 08-MAR-2002; 20C2WO-JP02213.
XX 09-MAR-2001; 20C1JP-0067282.
XX (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.
XX Aga H, Higashiyama T, Watanabe H, Sonoda T, Kubota M;
XX WPI; 2003-092811/08.
XX N-PSDB; ABV99948.
XX New cyclic tetrasaccharide compounds useful as sweetening agent for
XX foods, drinks or pharmaceuticals -
XX Disclosure; Page 122-129; 133pp; Japanese.
XX The present invention relates to novel cyclic tetrasaccharide compounds:
XX cyclo(-6-)-alpha-D-glucopyranosyl-(1-3)-alpha-D-glucopyranosyl-(1-6)-
XX alpha-D-glucopyranosyl-(1-3)-alpha-D-glucopyranosyl-(1-) derivatives.
XX The tetrasaccharide compounds are used as sweetening agents for foods,
XX drinks or pharmaceuticals. The present protein was used to illustrate the
XX invention.
XX SQ Sequence 1284 AA;

Query Match 100.0%; Score 43; DB 24; Length 1284;
Best Local Similarity 100.0%; Pred. No. 6.5; Mismatches 0; Indels 0; Gaps 0;
Matches 9; Conservative 0;

QY 1 YVSSLGNI 9
DB 36 YVSSLGNI 44
RESULT 7
ABP57724
ID ABP57724 standard; Protein; 1284 AA.
XX AC ABP57724;
XX DT 23-JAN-2003 (first entry)
XX DE Protein #1 related to isomaltose production.
XX KM Isomaltose.
XX OS Unidentified.
XX PN WO200288374-A1.
XX PD 07-NOV-2002.
XX PF 25-APR-2002; 2002WO-JP04166.
XX PR 27-APR-2001; 2001JP-0130922.
XX PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.
XX Kubota M, Nishimoto T, Higashiyama T, Watanabe H, Fukuda S;
XX Miyake T;
XX WPI; 2003-067722/06.
XX N-PSDB; ABV75608.
XX Production of isomaltose used for manufacturing e.g. beverages, health
XX foods and cosmetics by using specific non-reducing end
XX alpha-1,4-glucosyl linkage-containing saccharide and alpha-isomaltosyl
XX glucosaccharide synthase -
XX Disclosure; Page 103-111; 121pp; Japanese.
XX

Query Match 100.0%; Score 43; DB 24; Length 1284;
Best Local Similarity 100.0%; Pred. No. 6.5; Mismatches 0; Indels 0; Gaps 0;
Matches 9; Conservative 0;

QY 1 YVSSLGNI 9
DB 36 YVSSLGNI 44
RESULT 7
ABP57724
ID ABP57724 standard; Protein; 1284 AA.
XX AC ABP57724;
XX DT 23-JAN-2003 (first entry)
XX DE Protein #1 related to isomaltose production.
XX KM Isomaltose.
XX OS Unidentified.
XX PN WO200288374-A1.
XX PD 07-NOV-2002.
XX PF 25-APR-2002; 2002WO-JP04166.
XX PR 27-APR-2001; 2001JP-0130922.
XX PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.
XX Kubota M, Nishimoto T, Higashiyama T, Watanabe H, Fukuda S;
XX Miyake T;
XX WPI; 2003-067722/06.
XX N-PSDB; ABV75608.
XX Production of isomaltose used for manufacturing e.g. beverages, health
XX foods and cosmetics by using specific non-reducing end
XX alpha-1,4-glucosyl linkage-containing saccharide and alpha-isomaltosyl
XX glucosaccharide synthase -
XX Disclosure; Page 103-111; 121pp; Japanese.
XX

Query Match 100.0%; Score 43; DB 24; Length 1284;
Best Local Similarity 100.0%; Pred. No. 6.5; Mismatches 0; Indels 0; Gaps 0;
Matches 9; Conservative 0;

QY 1 YVSSLGNI 9
DB 36 YVSSLGNI 44
RESULT 7
ABP57724
ID ABP57724 standard; Protein; 1284 AA.
XX AC ABP57724;
XX DT 23-JAN-2003 (first entry)
XX DE Protein #1 related to isomaltose production.
XX KM Isomaltose.
XX OS Unidentified.
XX PN WO200288374-A1.
XX PD 07-NOV-2002.
XX PF 25-APR-2002; 2002WO-JP04166.
XX PR 27-APR-2001; 2001JP-0130922.
XX PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.
XX Kubota M, Nishimoto T, Higashiyama T, Watanabe H, Fukuda S;
XX Miyake T;
XX WPI; 2003-067722/06.
XX N-PSDB; ABV75608.
XX Production of isomaltose used for manufacturing e.g. beverages, health
XX foods and cosmetics by using specific non-reducing end
XX alpha-1,4-glucosyl linkage-containing saccharide and alpha-isomaltosyl
XX glucosaccharide synthase -
XX Disclosure; Page 103-111; 121pp; Japanese.
XX

Query Match 100.0%; Score 43; DB 24; Length 1284;
Best Local Similarity 100.0%; Pred. No. 6.5; Mismatches 0; Indels 0; Gaps 0;
Matches 9; Conservative 0;

QY 1 YVSSLGNI 9
DB 36 YVSSLGNI 44
RESULT 7
ABP57724
ID ABP57724 standard; Protein; 1284 AA.
XX AC ABP57724;
XX DT 23-JAN-2003 (first entry)
XX DE Protein #1 related to isomaltose production.
XX KM Isomaltose.
XX OS Unidentified.
XX PN WO200288374-A1.
XX PD 07-NOV-2002.
XX PF 25-APR-2002; 2002WO-JP04166.
XX PR 27-APR-2001; 2001JP-0130922.
XX PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.
XX Kubota M, Nishimoto T, Higashiyama T, Watanabe H, Fukuda S;
XX Miyake T;
XX WPI; 2003-067722/06.
XX N-PSDB; ABV75608.
XX Production of isomaltose used for manufacturing e.g. beverages, health
XX foods and cosmetics by using specific non-reducing end
XX alpha-1,4-glucosyl linkage-containing saccharide and alpha-isomaltosyl
XX glucosaccharide synthase -
XX Disclosure; Page 103-111; 121pp; Japanese.
XX

Query Match 100.0%; Score 43; DB 24; Length 1284;
Best Local Similarity 100.0%; Pred. No. 6.5; Mismatches 0; Indels 0; Gaps 0;
Matches 9; Conservative 0;

```

QY      1 YVSSLGNLI 9
Db      10 YVSTLGNAI 18

RESULT 9
AAB48343
ID      AAB48343 standard; Protein; 773 AA.
XX
AC      AAB48343;
DT      20-APR-2001 (first entry)
XX
DE      S. pneumoniae Sp130 polypeptide.
XX
KW      Immunogenic; Sp128; Sp130; pneumococcal; otitis media; nasopharyngeal;
KW      bronchial; lung; blood; infection; immune response; immunotherapy;
KW      antibacterial; auditory; vaccine.
XX
OS      Streptococcus pneumoniae.
XX
PN      WO200076540-A2.
XX
PD      21-DEC-2000.
XX
PF      09-JUN-2000; 2000WO-US15925.
XX
PR      10-JUN-1999; 99JS-0138453.
XX
PA      (MEDI-) MED IMMUNE INC.
XX
PI      Adamou JE, Choi GH;
XX
WPI: 2001-112197/12.
XX
N-PSDB; AAC84742.
XX
New vaccines comprising Sp128 or Sp130 polypeptides, for treating and
preventing pneumococcal infections, particularly infections caused by
Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or
blood infections
XX
Claim 8; Page 51-54; 54pp; English.
XX
The invention relates to novel immunogenic polypeptides, Sp128 and Sp130
from S. pneumoniae. Vaccines comprising the polypeptides are useful for
the treatment and prevention of pneumococcal infections, particularly
infections caused by Streptococcus, such as otitis media, nasopharyngeal,
bronchial, lung or blood infections. The antigens are used as immunogenic
agents to stimulate an immune response. The antisera and antibodies may
also be used in diagnosing and treating pneumococcal infections.
Recombinant polypeptides serve as a mechanism for stimulating production
of antibodies for use in passive immunotherapy, diagnostic reagents, and
as reagents in other processes such as affinity chromatography. The
present sequence represents the S. pneumoniae Sp130 polypeptide.
XX
SQ      Sequence 773 AA;
Query Match 81.4%; Score 35; DB 22; Length 773;
Best Local Similarity 77.8%; Pred. No. 1.6e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 YVSSLGNLI 9
Db      222 YVSSYGNGVI 230

RESULT 10
AAR80168
ID      AAR80168 standard; Protein; 831 AA.
XX
AC      AAR80168;
XX
DT      01-MAY-1996 (first entry)

pMIS1 MISP spider silk protein insert product.
Spider silk; repeat unit; consensus; minor ampullate silk protein;
spidroin; MISP; orb web spider; dragline.
Nephila clavipes.
Key Location/Qualifiers
Misc-difference 272..285
/note= "represented as indeterminate, corresponds
to a highly compressed GC rich region in
AAQ98470 which could not be sequenced"
WO9525165-A1.
21-SEP-1995.
14-MAR-1995; 95WO-US03139.
14-MAR-1994; 94US-0209747.
(UYWY-) UNIV WYOMING.
Colgin M, Lewis RV;
WPI: 1995-336970/43.
N-PSDB; AAQ98470.
Polypeptide(s) comprising repeated unit amino acid sequences, also
cDNAs - derived from minor ampullate spider silk proteins and used
to form spider silk fibres
Claim 2; Fig 1; 86pp; English.
This sequence represents the product of cDNA clone, pMIS1, encoding
the orb web spider minor ampullate silk protein MISP1 has been
identified and sequenced. Repeat unit peptides of MISP1 may form
part of a larger polypeptide with an amino terminus (see AAR80184-85)
and a carboxy terminus (see AAR80186-89). Repeat unit peptides make up
spider silk proteins (spidroins) which in turn aggregate to form the
silk fibres. Spider silk fibres have high tensile strength and
significant elasticity. An isolated cDNA clone of a silk protein
encoding sequence is of use to produce the protein at high yields
using recombinant DNA technology.
SQ      Sequence 831 AA;
Query Match 81.4%; Score 35; DB 16; Length 831;
Best Local Similarity 77.8%; Pred. No. 1.8e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 YVSSLGNLI 9
Db      11 YVSTLGNAI 19

RESULT 11
AAY81710
ID      AAY81710 standard; Protein; 2120 AA.
XX
AC      AAY81710;
XX
DT      02-JUN-2000 (first entry)
XX
DE      Streptococcus pneumoniae protein sequence ID3.
XX
KW      Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;
KW      bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;
KW      kidney disease; diabetes; immunosuppressive disorder; otitis media;
KW      pneumococcal septicaemia; sinusitis; meningitis; therapy.
XX
OS      Streptococcus pneumoniae.

```

Wed Jan 21 11:28:15 2004

us-10-089-549-1.rag

XX WO200006738-A2.
 XX 10-FEB-2000.
 XX 27-JUL-1999; 93WO-GB02452.
 XX 27-JUL-1998; 98GB-0316136.
 XX 19-MAR-1999; 99US-0125329.
 XX (MICR-) MICROBIAL TECHNIQS LTD.
 XX Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;
 XX WPI; 2000-195301/17.
 XX N-PSDB; AA291806.
 XX Streptococcal proteins and polynucleotides useful for diagnosis,
 XX treatment and prophylaxis of bacterial infections -
 XX Claim 2; Page 41-42; 76pp; English.
 XX This sequence represents a Streptococcus pneumoniae protein of the
 XX invention. The proteins (or their homologues, derivatives and/or
 XX fragments) are useful as immunogens or antigens. Immunogenic or antigenic
 XX compositions comprising the proteins are useful as vaccines and also in
 XX diagnostic assays. The sequences are useful for the detection or
 XX diagnosis of S. pneumoniae infection, by contacting a sample to be tested
 XX with them. Agents capable of antagonising, inhibiting or interfering with
 XX the function or expression of the protein or polypeptide are useful in
 XX medical compositions in the treatment or prophylaxis of S. pneumoniae
 XX infection. As the sequences can be used to treat S. pneumoniae infection,
 XX they can be used to treat bacterial pneumonia, which has high rates in
 XX young children, the elderly, and in patients with predisposing conditions
 XX such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,
 XX or with immunosuppressive disorders, especially AIDS. They can also be
 XX used to treat pneumococcal septicaemia, otitis media, sinusitis, and
 XX meningitis.
 XX SQ Sequence 2120 AA;
 Query Match 81.4%; Score 35; DB 21; Length 2120;
 Best Local Similarity 77.8%; Pred. No. 4.9e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 YVSSLGNI 9
 Db 1534 YVSSYGNVI 1542
 RESULT 12
 ABU01020
 ID ABU01020 standard; Protein; 2140 AA.
 XX AC ABU01020;
 XX 11-FEB-2003 (first entry)
 XX S. pneumoniae type 4 strain protein from coding region #590.
 XX Bacterial meningitis; pneumonia; sepsis; otitis media;
 XX ear infection; antiinflammatory; antibacterial; immunostimulant;
 XX auditory; respiratory; gene therapy; vaccine.
 XX Streptococcus pneumoniae type 4 strain.
 XX WO200277021-A2.
 XX 03-OCT-2002.
 XX 27-MAR-2002; 2002WO-1302163.
 XX 27-MAR-2001; 2001GB-0007658.

XX (CHIR-) CHIRON SPA.
 XX (GENO-) INST GENOMIC RES.
 XX Masignani V, Tettelin H, Fraser C;
 XX WPI; 2003-040579/03.
 XX N-PSDB; ABX06302.
 XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,
 XX useful as medicaments for treating or preventing a disease or infection
 XX due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
 XX or ear infection -
 XX Claim 1; SEQ ID No 1180; 56pp; English.
 XX The invention relates to a protein comprising or having at least 50%
 XX identity to any of the 2469 amino acid sequences, identified in the
 XX specification (available on a computer readable format), or its fragment,
 XX expressed from 2469 of 2489 identified DNA coding regions from the
 XX Streptococcus pneumoniae type 4 strain genomic sequence appearing as
 XX ABS56454. Also included are an antibody which binds one of the
 XX proteins, treating a patient by administering the protein, DNA or
 XX antibody (in a composition), a kit comprising first and second primers,
 XX which are the nucleic acid cited above or fragments between nucleotides
 XX 8-100 of a sequence not defined in the specification, for amplifying a
 XX target sequence contained within a Streptococcus nucleic acid sequence,
 XX where the first primer is substantially complementary to the target
 XX sequence and the second primer is substantially complementary to the
 XX sequence of the target sequence, and where the parts of the primers
 XX complement of the target sequence define the termini of the target
 XX having substantial complementarity define the termini of a test compound
 XX with the protein, and determining whether the test compound binds to the
 XX protein and a Streptococcus pneumoniae bacterium, where one or more
 XX genes encoding the proteins has been rendered inactive. The proteins,
 XX nucleic acid molecules, antibody and compositions are useful as
 XX medicaments for treating or preventing a disease or infection due to
 XX streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
 XX sepsis, otitis media or ear infection. They are also useful for
 XX vaccines, diagnostics and antibiotics. The methods are useful for
 XX identifying immunodominant proteins. The present sequence is one of
 XX the 2469 proteins expressed by the identified coding regions from the
 XX genomic sequence.
 XX Note: The sequence data for this patent did not form part
 XX of the printed specification, but was obtained in electronic
 XX format directly from WIPO at
 XX ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 2140 AA;
 Query Match 81.4%; Score 35; DB 24; Length 2140;
 Best Local Similarity 77.8%; Pred. No. 5e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 YVSSLGNI 9
 Db 1555 YVSSYGNVI 1563
 RESULT 13
 AA017340
 ID AA017340 standard; peptide; 9 AA.
 XX AC AA017340;
 XX 08-JUL-2002 (first entry)
 XX Alpha-isomaltosylglucosaccharide synthase related peptide #11.
 XX Alpha-isomaltosylglucosaccharide synthase; enzyme; sweetener; food;
 XX drink; cosmetics; pharmaceutical; cyclic tetrasaccharide.
 XX Bacillus globisporus.

XX WO200210361-A1.
 PN
 XX
 XX
 XX 07-FEB-2002.
 XX
 XX
 XX 25-JUL-2001; 2001WO-JP06412.
 PF
 XX
 XX 01-AUG-2000; 2000JP-0233364.
 PR
 XX 02-AUG-2000; 2000JP-0234937.
 PR
 XX (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.
 FA
 XX Kubota M, Tsusaki K, Higashiyama T, Fukuda S, Miyake T;
 PI
 XX WPI; 2002-315302/35.
 DR
 XX Alpha-Isomaltosylglucosaccharide synthase capable of transferring
 PT Alpha-glucosyl from saccharide, useful in producing cyclic
 PT tetrasaccharides and related carbohydrates industrially for application
 PT in compositions e.g. as drugs
 XX
 XX Claim 5; Page 203; 209pp; Japanese.
 PS
 XX The present invention relates to an alpha-isomaltosylglucosaccharide
 CC synthase which is capable of transferring alpha-glucosyl from a
 CC saccharide to form a specific saccharide which carries an alpha-1,6
 CC glucosyl linkage at its non-reducing end and an alpha-1,4 glucosyl
 CC linkage at ends other than the non-reducing end and has a degree of
 CC glucose polymerization of at most 3, but without substantially elevating
 CC the reducing ability. The synthase is useful in producing cyclic
 CC tetrasaccharides and related carbohydrates industrially for application
 CC in compositions e.g. as sweeteners, low calorie food materials, taste
 CC improvers, flavour improvers, quality improvers, water-separation
 CC inhibitors, stabilizers, excipients, binding agents and pulverisation
 CC bases, especially in foods, drinks, cosmetics and pharmaceuticals. The
 CC present sequence is a peptide described in the exemplification of the
 CC invention.
 XX
 XX Sequence 9 AA;
 SQ
 Query Match 76.7%; Score 33; DB 23; Length 9;
 Best Local Similarity 66.7%; Pred. No. 9.3e+05;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 YVSSLGNI 9
 DB 1 HVSAIGNLL 9
 :||:||||:
 :||:||||:
 RESULT 14
 ABG30538
 ID ABG30538 standard; Protein; 1251 AA.
 XX
 XX
 AC ABG30538;
 XX
 XX 07-OCT-2002 (first entry)
 DT
 XX Alpha-isomaltosylglucosaccharide synthase #2 mature protein.
 XX
 XX Alpha-isomaltosylglucosaccharide synthase; sugar production;
 KW cyclic tetrasaccharide; cyclo(-6)-alpha-D-glucopyranosyl(1-3);
 KW alpha-isomaltosyltransferase; viscosity modifier; sweetener;
 KW heat stabiliser; hydrolysis stabiliser; colour stabiliser; foodstuff;
 KW cosmetic; drug composition; enzyme.
 XX
 XX Unidentified.
 OS
 XX WO200255708-A1.
 XX
 XX 18-JUL-2002.
 PD
 XX 09-JAN-2002; 2002WO-JP00052.
 PF
 XX 12-JAN-2001; 2001JP-0005441.
 PR
 XX (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.
 PA
 XX Kubota M, Maruta K, Yamamoto T, Fukuda S;
 PI
 XX WPI; 2002-520129/55.
 DR
 XX N-PSDB; ABK88158.
 DR
 XX New alpha-isomaltosylglucosaccharide synthase of bacterial origin for

PR 12-JAN-2001; 2001JP-0005441.
 XX (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.
 PA Kubota M, Maruta K, Yamamoto T, Fukuda S;
 PI WPI; 2002-520129/55.
 DR N-PSDB; ABK88155.
 DR New alpha-isomaltosylglucosaccharide synthase of bacterial origin for
 PT the production of cyclic tetrasaccharide gum
 XX
 XX Claim 1; Page 105-109; 144pp; Japanese.
 PS
 XX The invention describes novel microbial polypeptides having
 CC alpha-isomaltosylglucosaccharide synthase activity. The proteins are
 CC useful for producing a sugar (I) having at least three glucose units
 CC connected by alpha-1,4 bonds except for alpha-1,6 at the non-reducing
 CC end. The invention also describes a method for the production of the
 CC cyclic tetrasaccharide cyclo(-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-
 CC glucopyranosyl(1-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-glucopyranosyl
 CC (I) by treatment of (I) with alpha-isomaltosyltransferase.
 CC (II) and similar sugars in crystalline or syrup form are used as
 CC a viscosity modifier, sweetener, heat stabiliser, hydrolysis stabiliser
 CC and colour stabiliser for foodstuffs, cosmetics and drug compositions.
 CC This is the amino acid sequence of an alpha-isomaltosylglucosaccharide
 CC synthase mature protein.
 XX
 XX Sequence 1251 AA;
 SQ
 Query Match 76.7%; Score 33; DB 23; Length 1251;
 Best Local Similarity 66.7%; Pred. No. 7.1e+02;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 YVSSLGNI 9
 DB 1 HVSAIGNLL 9
 :||:||||:
 :||:||||:
 RESULT 15
 ABG30564
 ID ABG30564 standard; Protein; 1286 AA.
 XX
 XX
 AC ABG30564;
 XX
 XX 07-OCT-2002 (first entry)
 DT
 XX Alpha-isomaltosylglucosaccharide synthase #2.
 XX
 XX Alpha-isomaltosylglucosaccharide synthase; sugar production;
 KW cyclic tetrasaccharide; cyclo(-6)-alpha-D-glucopyranosyl(1-3);
 KW alpha-isomaltosyltransferase; viscosity modifier; sweetener;
 KW heat stabiliser; hydrolysis stabiliser; colour stabiliser; foodstuff;
 KW cosmetic; drug composition.
 XX
 XX Unidentified.
 OS
 XX WO200255708-A1.
 XX
 XX 18-JUL-2002.
 PD
 XX 09-JAN-2002; 2002WO-JP00052.
 PF
 XX 12-JAN-2001; 2001JP-0005441.
 PR
 XX (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.
 PA
 XX Kubota M, Maruta K, Yamamoto T, Fukuda S;
 PI
 XX WPI; 2002-520129/55.
 DR
 XX N-PSDB; ABK88158.
 DR
 XX New alpha-isomaltosylglucosaccharide synthase of bacterial origin for

PT the production of cyclic tetrasaccharide gpm -
 XX Disclosure; Page 128-134; 144pp; Japanese.
 XX
 CC The invention describes novel microbial polypeptides having
 CC alpha-isomaltosylglucosaccharide synthase activity. The proteins are
 CC useful for producing a sugar (I) having at least three glucose units
 CC connected by alpha-1,4 bonds except for alpha-1,6 at the non-reducing
 CC end. The invention also describes a method for the production of the
 CC cyclic tetrasaccharide cyclo(-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-
 CC glucopyranosyl(-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-glucopyranosyl
 CC (1-) (II) by treatment of (I) with alpha-isomaltosyltransferase.
 CC (II) and similar sugars in crystalline or syrup form are used as
 CC a viscosity modifier, sweetener, heat stabiliser, hydrolysis stabiliser
 CC and colour stabiliser for foodstuffs, cosmetics and drug compositions.
 CC This is the amino acid sequence of an alpha-isomaltosylglucosaccharide
 CC synthase.
 XX
 SQ Sequence 1286 AA;

Query Match 76.7%; Score 33; DB 23; Length 1286;
 Best Local Similarity 66.7%; Pred No: 7.3e+02;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 YVSSLGNLI 9
 Db :||:||||:
 36 HVSAIGNLL 44

Search completed: January 20, 2004, 16:52:37
 Job time : 34.8015 secs

Sequence 4, Appli
Sequence 34, Appli
Sequence 34, Appli
Sequence 34, Appli
Sequence 34, Appli
Sequence 34, Appli
Sequence 304, App
Sequence 302, App
Sequence 4196, Ap
Sequence 1112, Ap
Sequence 28607, A
Sequence 12, Appli
Sequence 149, App
Sequence 151, App
Sequence 149, App
Sequence 150, App
Sequence 151, App

74.4 1179 3 US-08-855-160-4
74.4 1182 1 US-08-349-867-34
74.4 1182 2 US-08-598-305A-34
74.4 1188 1 US-08-239-476-34
74.4 1188 2 US-08-639-923A-34
74.4 1188 5 PCT-US95-05431-34
72.1 244 4 US-09-071-035-304
72.1 269 4 US-09-071-035-302
72.1 297 4 US-09-107-532A-4196
72.1 376 4 US-09-198-452A-1112
72.1 86C 4 US-09-252-991A-28607
72.1 1698 3 US-09-315-793-12
69.8 28 3 US-08-486-099-149
69.8 28 3 US-08-486-099-150
69.8 28 3 US-08-486-099-151
69.8 28 3 US-08-484-223B-149
69.8 28 3 US-08-484-223B-150
69.8 28 3 US-08-484-223B-151

ALIGNMENTS

RESULT 1
US-08-209-747-15
; Sequence 15, Application US/08209747
; Patent No. 5733771
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; TITLE OF INVENTION: cDNAs Encoding Minor Ampullate Spider
; TITLE OF INVENTION: Silk Proteins
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-3487
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/209,747
; FILING DATE: 14-MAR-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-104P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 155 amino acids
; TYPE: amino acid
; STRANDBESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: N-terminal
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..155
; OTHER INFORMATION: /label= MISPN aa
; OTHER INFORMATION: /note= amino-terminal sequence of mispl, see Fig.
; OTHER INFORMATION: 4*
; OTHER INFORMATION:
US-08-209-747-15

Query Match 81.4%; Score 35; DB 1; Length 155;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:53:13 ; Search time 10.5802 Seconds
(without alignments)
35.992 Million cell updates/sec

Title: US-10-089-549-1

Perfect score: 43

Sequence: 1 YVSSLGNLI 9

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCBUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	81.4	155	1	US-08-209-747-15
2	35	81.4	155	1	US-08-458-298-15
3	35	81.4	832	1	US-08-209-747-2
4	35	81.4	832	1	US-08-458-298-2
5	33	76.7	572	4	US-09-648-004-8
6	33	76.7	622	4	US-09-328-352-7970
7	32	74.4	334	4	US-09-252-991A-18454
8	32	74.4	419	3	US-09-554-225-1
9	32	74.4	490	4	US-09-328-352-4607
10	32	74.4	595	6	5523211-3
11	32	74.4	613	2	US-08-622-740-6
12	32	74.4	613	3	US-08-440-689-6
13	32	74.4	613	4	US-09-122-399-6
14	32	74.4	613	4	US-08-447-985-11
15	32	74.4	617	3	US-08-729-601A-2
16	32	74.4	898	3	US-08-863-102-1
17	32	74.4	898	3	US-08-863-102-2
18	32	74.4	907	3	US-08-863-102-4
19	32	74.4	969	1	US-07-671-817A-4
20	32	74.4	1156	3	US-09-178-252-15
21	32	74.4	1177	1	US-07-828-788A-8
22	32	74.4	1177	5	PCT-US92-11337-8
23	32	74.4	1177	6	5169629-2
24	32	74.4	1178	1	US-08-446-486-5
25	32	74.4	1178	1	US-08-463-308-5
26	32	74.4	1178	1	US-08-463-308-5
27	32	74.4	1178	6	5254799-5

[illegible]

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,298
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/203,747
; FILING DATE: 14-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-104P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 832 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: N. clavipes
; TISSUE TYPE: minor ampullate gland
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..309
; US-08-458-288-2

Query Match 81.4%; Score 35; DB 1; Length 832;
Best Local Similarity 77.8%; Pred. No. 1e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

QY 1 YVSSLGNI 9
Db 1 YVSTLGNAI 19

RESULT 5
US-09-648-004-8
; Sequence 8, Application US/09648004
; Patent No. 6498242
; GENERAL INFORMATION:
; APPLICANT: CHEN, QIONG
; APPLICANT: THOMAS, STUART
; APPLICANT: NAGARAJAN, VASANTHA
; TITLE OF INVENTION: BIOLOGICAL METHOD FOR THE PRODUCTION OF ADIPIC ACID AND
; TITLE OF INVENTION: INTERMEDIATES
; FILE REFERENCE: CL-1341-A
; CURRENT APPLICATION NUMBER: US/09/648,004
; CURRENT FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/252,553
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 572
; TYPE: PRT
; ORGANISM: Acinetobacter sp.
; US-09-648-004-8

Query Match 76.7%; Score 33; DB 4; Length 572;
Best Local Similarity 66.7%; Pred. No. 1.7e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 1; Gaps 0;

QY 1 YVSSLGNI 9
Db 1 YVSTLGNAI 19

us-10-089-549-1.ra1

Db 119 YVISLGNV 127

RESULT 6
US-09-328-352-7970
; Sequence 7970, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7970
; LENGTH: 622
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
; US-09-328-352-7970

Query Match 76.7%; Score 33; DB 4; Length 622;
Best Local Similarity 66.7%; Pred. No. 1.8e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 1; Gaps 0;

QY 1 YVSSLGNI 9
Db 169 YVISLGNV 177

RESULT 7
US-09-252-991A-18454
; Sequence 18454, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18454
; LENGTH: 334
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-18454

Query Match 74.4%; Score 32; DB 4; Length 334;
Best Local Similarity 75.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

QY 1 YVSSLGNI 8
Db 120 YIISLGNL 127

RESULT 8
US-09-554-225-1
; Sequence 1, Application US/09554225
; Patent No. 6252056
; GENERAL INFORMATION:
; APPLICANT: FUKUSHIMA, DAIKICHI
; APPLICANT: NAKADE, SHINJI
; APPLICANT: HAGA, HISANORI
; TITLE OF INVENTION: HUMAN LYSOPHOSPHATIDIC ACID RECEPTOR AND USE THEREOF
; FILE REFERENCE: Q59116
; CURRENT APPLICATION NUMBER: US/09/554,225
; CURRENT FILING DATE: 2000-05-11
```

us-10-089-549-1.ra1

Wed Jan 21 11:28:15 2004

```

; PRIOR APPLICATION NUMBER: PCT/JP98/05047
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: JP P.Hei 9-307749
; PRIOR FILING DATE: 1997-11-11
; NUMBER OF SEQ ID NOS: 15
; SEQ ID NO 1
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-09-554-225-1

Query Match 74.4%; Score 32; DB 3; Length 419;
Best Local Similarity 66.7%; Pred. No. 1.9e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 YVSSLGNLI 9
Db 93 FVSPFLGNLV 101

RESULT 9
US-09-328-352-4607
; Sequence 4607, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Bretor et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03FA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4607
; LENGTH: 490
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4607

Query Match 74.4%; Score 32; DB 4; Length 490;
Best Local Similarity 66.7%; Pred. No. 2.2e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 YVSSLGNLI 9
Db 54 YVDEVLNLI 62

RESULT 10
5523211-3
; Patent No. 5523211
; APPLICANT: PUSZTAI-CAREY, MARIANNE; CAREY, PAUL R.; LESSARD,
; TIMOTHY; YAGUCHI, MAKOTO
; TITLE OF INVENTION: IDENTIFICATION, QUANTITATION AND
; PURIFICATION OF INSECTICIDAL PROTEINS FROM BACILLUS THURINGIENSIS
; NUMBER OF SEQUENCES: 3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/277,076
; FILING DATE: 19-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 102,491
; FILING DATE: 05-AUG-1993
; APPLICATION NUMBER: 836,967
; FILING DATE: 19-FEB-1992
; APPLICATION NUMBER: 493,453
; FILING DATE: 14-MAR-1990
; SEQ ID NO: 3
; LENGTH: 595
5523211-3

Query Match 74.4%; Score 32; DB 6; Length 595;
Best Local Similarity 55.6%; Pred. No. 2.7e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

; PRIOR APPLICATION NUMBER: PCT/JP98/05047
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: JP P.Hei 9-307749
; PRIOR FILING DATE: 1997-11-11
; NUMBER OF SEQ ID NOS: 15
; SEQ ID NO 1
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-08-622-740-6

Query Match 74.4%; Score 32; DB 2; Length 613;
Best Local Similarity 55.6%; Pred. No. 2.8e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 YVSSLGNLI 9
Db 579 FTSSLGNIV 587

RESULT 12
US-08-440-689-6
; Sequence 6, Application US/08440689
; Patent No. 6025545
; GENERAL INFORMATION:
; APPLICANT: Lundquist, Ronald C.
; APPLICANT: Walters, David A.
; APPLICANT: Kirihara, Julie A.
; TITLE OF INVENTION: Methods and Compositions for the
; TITLE OF INVENTION: Production of Stably Transformed, Fertile Monocot Plants
; TITLE OF INVENTION: and Cells Thereof
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.
; STREET: 3500 IDS Center
; CITY: Minneapolis

```

STATE: MN
COUNTRY: USA
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,689
FILING DATE: 15-MAY-1995
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Woessner, Warren D.
REGISTRATION NUMBER: 30,440
REFERENCE/DOCKET NUMBER: 950.013US2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-339-0331
TELEFAX: 612-339-3061
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 613 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-440-689-6

Query Match 74.4%; Score 32; DB 3; Length 613;
Best Local Similarity 55.6%; Pred. No. 2.8e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 YVSSLGNI 9
DB 579 FTSSLGNI 587

RESULT 13
US-09-122-399-6
Sequence 6, Application US/09122399
Patent No. 6329574
GENERAL INFORMATION:
APPLICANT: Lundquist, Ronald C.
APPLICANT: Walters, David A.
APPLICANT: Kirihara, Julie A.
TITLE OF INVENTION: Methods and Compositions for the
Production of Stably Transformed, Fertile Monocot Plants
TITLE OF INVENTION: and Cells Thereof
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSER: Schwegman, Lundberg & Woessner, P.A.
STREET: 3500 IDS Center
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/122,399
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/112,245
FILING DATE: 25-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Woessner, Warren D.
REGISTRATION NUMBER: 30,440
REFERENCE/DOCKET NUMBER: 950.13US01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-339-0331

TELEFAX: 612-339-3061
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 613 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-122-399-6

Query Match 74.4%; Score 32; DB 4; Length 613;
Best Local Similarity 55.6%; Pred. No. 2.8e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 YVSSLGNI 9
DB 579 FTSSLGNI 587

RESULT 14
US-08-447-985-11
Sequence 11, Application US/08447985
Patent No. 6399861
GENERAL INFORMATION:
APPLICANT: Adams, Thomas R. et al.
TITLE OF INVENTION: Methods and Compositions for the
Production of Stably Transformed, Fertile Monocot Plants
and Cells Thereof
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSER: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/447,985
FILING DATE: 23-May-1995
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/113,561
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: DEKM-055/PAR
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 713/789-2679
TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 613 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-08-447-985-11

Query Match 74.4%; Score 32; DB 4; Length 613;
Best Local Similarity 55.6%; Pred. No. 2.8e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 YVSSLGNI 9
DB 579 FTSSLGNI 587

Wed Jan 21 11:28:15 2004

```

RESULT 15
US-08-729-601A-2
; Sequence 2, Application US/08729601A
; Patent No. 6186302
; GENERAL INFORMATION:
; APPLICANT: Merlo, Donald J.
; APPLICANT: Folkerts, Otto
; TITLE OF INVENTION: Modified Bacillus Thuringiensis Gene for
; TITLE OF INVENTION: Lepidopteran Control in Plants
; NUMBER OF SEQUENCES: 84
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fitch, Even, Tabin & Flannery
; STREET: 135 S. LaSalle St.
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/729,601A
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Krueger, James P.
; REGISTRATION NUMBER: 35,234
; REFERENCE/DOCKET NUMBER: 80089
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-372-7842
; TELEFAX: 312-372-7848
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 617 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-729-601A-2

```

```

Query Match          74.4%; Score 32; DB 3; Length 617;
Best Local Similarity 55.6%; Pred. No. 2.8e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 YVSSIGNLI 9
Db      579 FTSSIGNIV 587

```

```

Search completed: January 20, 2004, 17:47:04
Job time : 11.5802 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: January 20, 2004, 16:54:27 ; Search time 23.8397 Seconds
(without alignments)
77.196 Million cell updates/sec

Title: US-10-089-549-1

Perfect score: 43

Sequence: 1 YVSSLGNI 9

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 762491 seqs, 204481190 residues

Total number of hits satisfying chosen parameters: 762491

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

```
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	100.0	9	12	US-10-089-549-1
2	35	81.4	334	12	US-10-369-493-19327
3	35	81.4	773	14	US-10-067-385-8
4	35	81.4	2119	12	US-09-769-744A-28
5	33	76.7	9	12	US-10-089-549-11
6	33	76.7	407	12	US-10-369-493-3506
7	33	76.7	572	15	US-10-272-419-8
8	32	74.4	188	9	US-09-864-761-34258
9	32	74.4	271	12	US-10-108-260A-4309
10	32	74.4	312	14	US-10-004-717-64
11	32	74.4	332	12	US-10-094-749-2426
12	32	74.4	369	9	US-09-826-508-36
13	32	74.4	374	12	US-10-032-585-7790
14	32	74.4	419	9	US-09-826-508-34
15	32	74.4	419	11	US-09-811-838-8
					Sequence 1, Appli
					Sequence 19327, A
					Sequence 8, Appli
					Sequence 28, Appl
					Sequence 11, Appl
					Sequence 3506, Ap
					Sequence 8, Appli
					Sequence 34258, A
					Sequence 4309, Ap
					Sequence 64, Appl
					Sequence 2426, Ap
					Sequence 36, Appl
					Sequence 7790, Ap
					Sequence 34, Appl
					Sequence 8, Appli

Query Match 100.0%; Score 43; DB 12; Length 9;

Best Local Similarity 100.0%; Pred No. 6.7e+05; Indels 0; Gaps 0;

Matches 9; Conservative 0; Mismatches 0;

QY 1 YVSSLGNI 9

Db 1 YVSSLGNI 9

16	32	74.4	419	12	US-09-953-686-8	Sequence 8, Appli
17	32	74.4	419	12	US-10-272-983-2	Sequence 2, Appli
18	32	74.4	419	12	US-10-375-703-1	Sequence 1, Appli
19	32	74.4	419	12	US-10-393-807-2	Sequence 2, Appli
20	32	74.4	419	12	US-10-417-820A-2	Sequence 2, Appli
21	32	74.4	419	12	US-10-455-127-2	Sequence 2, Appli
22	32	74.4	419	15	US-10-225-567A-569	Sequence 569, App
23	32	74.4	589	12	US-10-102-469-6	Sequence 6, Appli
24	32	74.4	640	12	US-10-102-469-4	Sequence 4, Appli
25	32	74.4	973	12	US-10-369-493-21064	Sequence 21064, A
26	32	74.4	1156	9	US-09-826-660-15	Sequence 15, Appl
27	32	74.4	1178	9	US-09-851-194-2	Sequence 2, Appli
28	32	74.4	1178	12	US-10-102-469-10	Sequence 10, Appl
29	32	74.4	1178	12	US-10-102-469-12	Sequence 12, Appl
30	32	74.4	1178	12	US-10-102-469-14	Sequence 14, Appl
31	32	74.4	1179	15	US-10-035-060-4	Sequence 4, Appli
32	31	72.1	108	10	US-09-764-846-169	Sequence 169, App
33	31	72.1	108	15	US-10-091-483-169	Sequence 169, App
34	31	72.1	193	12	US-10-230-331-23	Sequence 23, Appl
35	31	72.1	193	12	US-10-310-154-663	Sequence 663, App
36	31	72.1	255	12	US-10-017-161-2180	Sequence 2180, Ap
37	31	72.1	255	12	US-10-292-798-1826	Sequence 1826, Ap
38	31	72.1	376	12	US-10-289-762-1112	Sequence 1112, Ap
39	31	72.1	543	12	US-10-369-493-21190	Sequence 21190, A
40	31	72.1	1486	9	US-09-815-242-13728	Sequence 13728, A
41	31	72.1	1517	9	US-09-815-242-10325	Sequence 10325, A
42	30.5	70.9	303	11	US-09-832-522-81	Sequence 81, Appl
43	30.5	70.9	312	10	US-09-886-055-213	Sequence 213, App
44	30.5	70.9	312	11	US-09-804-291-213	Sequence 213, App
45	30.5	70.9	312	11	US-09-832-522-24	Sequence 24, Appl

ALIGNMENTS

RESULT 1

```
US-10-089-549-1
; Sequence 1, Application US/10089549
; Publication No. US20030194762A1
; GENERAL INFORMATION:
; APPLICANT: KUBOTA, Michio
; APPLICANT: TSUTSUKI, Keiji
; APPLICANT: HIGASHIYAMA, Takanobu
; APPLICANT: FUKUDA, Shigeharu
; APPLICANT: MIYAKE, Toshio
; TITLE OF INVENTION: ALPHA-ISOMALTOSEGLUCOSACCHARIDE-FORMING ENZYME, PROCESS AND U
; FILE REFERENCE: KUBOTA-9
; CURRENT APPLICATION NUMBER: US/10/089,549
; CURRENT FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: JP 233364/2000
; PRIOR FILING DATE: 2000-08-01
; PRIOR APPLICATION NUMBER: JP 234937/2000
; PRIOR FILING DATE: 2000-08-02
; PRIOR APPLICATION NUMBER: PCT/JP01/06412
; PRIOR FILING DATE: 2001-07-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Bacillus globisporus
US-10-089-549-1
```

RESULT 2
 US-10-369-493-19327
 ; Sequence 19327, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 19327
 ; LENGTH: 334
 ; TYPE: PRT
 ; ORGANISM: Myxococcus xanthus
 US-10-369-493-19327

Query Match 81.4%; Score 35; DB 12; Length 334;
 Best Local Similarity 66.7%; Pred. No. 64;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YVSSLGNI 9
 |||||
 Db 154 YVSQGNLV 162

RESULT 3
 US-10-067-385-8
 ; Sequence 8, Application US/10067385
 ; Publication No. US20020110562A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Choi, Gil
 ; APPLICANT: Adamo, John
 ; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines
 ; FILE REFERENCE: 469201-589
 ; CURRENT APPLICATION NUMBER: US/10/067,385
 ; CURRENT FILING DATE: 2002-02-05
 ; PRIOR APPLICATION NUMBER: US/09/590,991
 ; PRIOR FILING DATE: 2000-06-09
 ; PRIOR APPLICATION NUMBER: US/60/138,453
 ; PRIOR FILING DATE: 1999-06-10
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 8
 ; LENGTH: 773
 ; TYPE: PRT
 ; ORGANISM: Streptococcus pneumoniae
 US-10-067-385-8

Query Match 81.4%; Score 35; DB 14; Length 773;
 Best Local Similarity 77.8%; Pred. No. 1.6e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YVSSLGNI 9
 |||||
 Db 222 YVSQGNLV 230

RESULT 4
 US-09-769-744A-28
 ; Sequence 28, Application US/09769744A
 ; Publication No. US20030134407A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Le Page, Richard WF
 ; APPLICANT: Wells, Jeremy M
 ; APPLICANT: Hanniffy, Sean B

; APPLICANT: Hansbro, Philip M
 ; TITLE OF INVENTION: Proteins
 ; FILE REFERENCE: PWC/P21122WO
 ; CURRENT APPLICATION NUMBER: US/09/769,744A
 ; CURRENT FILING DATE: 2001-01-26
 ; PRIOR APPLICATION NUMBER: PCT/GB99/02452
 ; PRIOR FILING DATE: 1999-07-27
 ; PRIOR APPLICATION NUMBER: GB 9816336.3
 ; PRIOR FILING DATE: 1998-07-27
 ; PRIOR APPLICATION NUMBER: US 60/125329
 ; PRIOR FILING DATE: 1999-03-19
 ; NUMBER OF SEQ ID NOS: 196
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 28
 ; LENGTH: 2119
 ; TYPE: PRT
 ; ORGANISM: Streptococcus pneumoniae
 US-09-769-744A-28

Query Match 81.4%; Score 35; DB 12; Length 2119;
 Best Local Similarity 77.8%; Pred. No. 4.8e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YVSSLGNI 9
 |||||
 Db 1534 YVSQGNLV 1542

RESULT 5
 US-10-089-549-11
 ; Sequence 11, Application US/10089549
 ; Publication No. US20030194762A1
 ; GENERAL INFORMATION:
 ; APPLICANT: KUBOTA, Michio
 ; APPLICANT: TSUSAKI, Keiji
 ; APPLICANT: HIGASHIYAMA, Takarobu
 ; APPLICANT: FUKUDA, Shigeharu
 ; APPLICANT: MIYAKE, Toshio
 ; TITLE OF INVENTION: ALPHA-ISONALTOSYLGLUCOSACCHARIDE-FORMING ENZYME, PROCESS AND
 ; FILE REFERENCE: KUBOTA=9
 ; CURRENT APPLICATION NUMBER: US/10/089,549
 ; CURRENT FILING DATE: 2002-12-13
 ; PRIOR APPLICATION NUMBER: JP 233364/2000
 ; PRIOR FILING DATE: 2000-08-01
 ; PRIOR APPLICATION NUMBER: JP 234937/2000
 ; PRIOR FILING DATE: 2000-08-02
 ; PRIOR APPLICATION NUMBER: PCT/JP01/06412
 ; PRIOR FILING DATE: 2001-07-25
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 11
 ; LENGTH: 9
 ; TYPE: PRT
 ; ORGANISM: Bacillus globisporus
 US-10-089-549-11

Query Match 76.7%; Score 33; DB 12; Length 9;
 Best Local Similarity 66.7%; Pred. No. 6.7e+05;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YVSSLGNI 9
 :|||:
 Db 1 HVSSALGNLI 9

RESULT 6
 US-10-369-493-3506
 ; Sequence 3506, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.


```

; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 3506
; LENGTH: 407
; TYPE: PRT
; ORGANISM: Neurospora crassa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(407)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-3506

Query Match          76.7%; Score 33; DB 12; Length 407;
Best Local Similarity 75.0%; Pred. No. 2e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      2 VSSIGNLI 9
Db      127 ISSIGNEL 134

RESULT 7
US-10-272-419-8
; Sequence 8, Application US/10272419
; Publication No. US20030087403A1
; GENERAL INFORMATION:
; APPLICANT: CHEN, QIONG
; APPLICANT: THOMAS, STUART
; APPLICANT: NAGARAJAN, VASANTHA
; TITLE OF INVENTION: BIOLOGICAL METHOD FOR THE PRODUCTION OF ADIPIC ACID AND
; TITLE OF INVENTION: INTERMEDIATES
; FILE REFERENCE: CL1341-A
; CURRENT APPLICATION NUMBER: US/10/272,419
; CURRENT FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: 09/252,553
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 572
; TYPE: PRT
; ORGANISM: Acinetobacter sp.
US-10-272-419-8

Query Match          76.7%; Score 33; DB 15; Length 572;
Best Local Similarity 66.7%; Pred. No. 2.9e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      1 YVSSIGNLI 9
Db      119 YVISIGNLV 127

RESULT 8
US-09-864-761-34258
; Sequence 34258, Application US/0986476;
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

```

```

; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34258
; LENGTH: 188
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURES:
; OTHER INFORMATION: MAP TO AL033379.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.9
; OTHER INFORMATION: EST_HUMAN HIT: BE798784.1, EVALUE 1.00e-11
; OTHER INFORMATION: SWISSPROT HIT: Q91081, EVALUE 6.00e-17
US-09-864-761-34258

Query Match          74.4%; Score 32; DB 9; Length 188;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      1 YVSSIGNLI 9
Db      43 FVSFLGNLV 51

RESULT 9
US-10-108-260A-4309
; Sequence 4309, Application US/10108260A

```

```

; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOKIHO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOMUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2426
; LENGTH: 332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2426

Query Match      74.4%; Score 32; DB 12; Length 332;
Best Local Similarity 66.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 YVSSLGNLI 9
DB 225 YFSGLNII 233

RESULT 12
US-09-826-508-36
; Sequence 36, Application US/09826508
; Patent No. US20010025093A1
; GENERAL INFORMATION:
; APPLICANT: Nabil Elshourbagy
; APPLICANT: Lisa Vawter
; TITLE OF INVENTION: G Protein-Coupled Receptor Polypeptides
; TITLE OF INVENTION: and Polynucleotides
; FILE REFERENCE: GP-70744USB
; CURRENT APPLICATION NUMBER: US/09/826,508
; CURRENT FILING DATE: 2001-04-05
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36
; LENGTH: 369
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-826-508-36

Query Match      74.4%; Score 32; DB 9; Length 369;
Best Local Similarity 66.7%; Pred. No. 2.8e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 YVSSLGNLI 9
DB 43 FVSFLGNLV 51

RESULT 13
US-10-032-585-7790
; Sequence 7790, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey

```

; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7790
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7790

Query Match 74.4%; Score 32; DB 12; Length 374;

Best Local Similarity 55.6%; Pred. No. 2.9e+02;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 YVSSIGNLI 9

||| |||:

Db 101 YVEALGNVV 109

RESULT 14

US-09-826-508-34

; Sequence 34, Application US/09826508

; Patent No. US20010025099A1

; GENERAL INFORMATION:

; APPLICANT: Nabil Elshourbagy

; APPLICANT: Lisa Vawter

; TITLE OF INVENTION: G Protein-Coupled Receptor Polypeptides

; TITLE OF INVENTION: and Polynucleotides

; FILE REFERENCE: GP-70744USE

; CURRENT APPLICATION NUMBER: US/09/826,508

; CURRENT FILING DATE: 2001-04-05

; NUMBER OF SEQ ID NOS: 40

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 34

; LENGTH: 419

; TYPE: PRT

; ORGANISM: HOMO SAPIENS

US-09-826-508-34

Query Match 74.4%; Score 32; DB 9; Length 419;

Best Local Similarity 66.7%; Pred. No. 3.3e+02;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 YVSSIGNLI 9

||| |||:

Db 93 FVSPFLGNLV 101

RESULT 15

US-09-811-838-8

; Sequence 8, Application US/09811838

; Publication No. US20030027800A1

; GENERAL INFORMATION:

; APPLICANT: Miller, Duane D.

; APPLICANT: Tigyi, Gabor

; APPLICANT: Dalton, James T.

; APPLICANT: Sardar, Vineet M.

; APPLICANT: Elrod, Don B.

; APPLICANT: Xu, Huiping

; APPLICANT: Baker, Daniel L.

; APPLICANT: Wang, Dean

; APPLICANT: Liliom, Karoly

; APPLICANT: Fischer, David J.

; APPLICANT: Virag, Tamas

; APPLICANT: Nusser, No. US20030027800A1a

; TITLE OF INVENTION: LPA RECEPTOR AGONISTS AND ANTAGONISTS AND METHODS OF

; TITLE OF INVENTION: USE

; FILE REFERENCE: 20609/181

; CURRENT APPLICATION NUMBER: US/09/811,838

; CURRENT FILING DATE: 2001-03-19

; PRIOR APPLICATION NUMBER: 60/190,370

; PRIOR FILING DATE: 2000-03-17

; NUMBER OF SEQ ID NOS: 26

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 8

; LENGTH: 419

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-811-838-8

Query Match 74.4%; Score 32; DB 11; Length 419;

Best Local Similarity 66.7%; Pred. No. 3.3e+02;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 YVSSIGNLI 9

||| |||:

Db 93 FVSPFLGNLV 101

Search completed: January 20, 2004, 17:53:05

Job time : 25.8397 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:43:26 ; Search time 9.34351 Seconds
(without alignments)
92.633 Million cell updates/sec

Title: US-10-089-549-1

Perfect score: 43

Sequence: 1 YVSSLGNI 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76:*

1: Pirl:*

2: Pirl2:*

3: Pirl3:*

4: Pirl4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	81.4	148	2 F71207	hypothetical prote
2	35	81.4	236	2 S42069	TEGT protein - rat
3	35	81.4	447	2 S28386	gene A protein - y
4	35	81.4	2140	2 F95374	serine proteinase,
5	35	81.4	2144	2 A97342	metalloproteinase,
6	33	76.7	153	2 C72420	hypothetical prote
7	33	76.7	220	2 T43580	probable oligomyci
8	33	76.7	512	2 T11154	cytochrome-c oxida
9	33	76.7	535	2 S58740	cytochrome-c oxida
10	33	76.7	2139	2 S46404	vitellogenin - yel
11	32	74.4	52	2 S43056	hypothetical prote
12	32	74.4	207	2 I53154	scleraxis - mouse
13	32	74.4	208	2 C64347	hypothetical prote
14	32	74.4	282	2 AD0513	bis(5'-nucleosyl)-
15	32	74.4	312	2 A40708	basic-helix-loop-h
16	32	74.4	334	2 G72332	hypothetical prote
17	32	74.4	348	2 B84448	hypothetical prote
18	32	74.4	356	2 G81907	probable integral
19	32	74.4	356	2 B81107	transporter, proba
20	32	74.4	551	2 D83277	electron transfer
21	32	74.4	579	1 W2B334	gene 34 protein -
22	32	74.4	618	2 S11445	parasporal crystal
23	32	74.4	730	2 G75292	excinuclease ABC c
24	32	74.4	761	2 JC7821	beta-N-acetylhexos
25	32	74.4	822	2 S68210	chloride channel p
26	32	74.4	907	2 S23399	chloride channel p
27	32	74.4	1073	2 T38763	hypothetical prote
28	32	74.4	1177	2 A49785	parasporal crystal
29	32	74.4	1178	1 US85XH	parasporal crystal

30	32	74.4	2150	2 T32497	hypothetical prote
31	32	74.4	2364	2 I40864	Cytotoxin L - Clo
32	31	72.1	180	2 D83623	hypothetical prote
33	31	72.1	193	1 DSECF	superoxide dismuta
34	31	72.1	193	2 E90924	superoxide dismuta
35	31	72.1	193	2 A85773	superoxide dismuta
36	31	72.1	193	2 AG0695	superoxide dismuta
37	31	72.1	213	2 E64568	superoxide dismuta
38	31	72.1	213	2 E71861	superoxide dismuta
39	31	72.1	220	2 S47166	superoxide dismuta
40	31	72.1	220	2 H81434	superoxide dismuta
41	31	72.1	243	2 B84864	hypothetical prote
42	31	72.1	269	2 E90309	conserved hypothet
43	31	72.1	302	2 F90162	hypothetical prote
44	31	72.1	311	2 E71689	hypothetical prote
45	31	72.1	312	2 T35940	probable transport

ALIGNMENTS

RESULT 1

F71207
hypothetical protein PH1926 - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
C;Accession: F71207
R;Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophili
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: F71207
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-148 <XAW>
A;Cross-references: GB:AP000007; MID:g3236134; PIDN:BAA31053.1; PID:g3258370
A;Experimental source: strain Orl
A;Note: this accession replaces an interim accession for a sequence replaced by GenBa
C;Genetics:
A;Gene: PH1926

Query Match 81.4%; Score 35; DB 2; Length 148;
Best Local Similarity 55.6%; Pred. No. 7.8;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 YVSSLGNI 9
| : | : | : | :
DB 100 YIDSIGNLV 108

RESULT 2

S42069
TEGT protein - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 21-Jan-2000
C;Accession: S42069; I57015; I76675
R;Guenther, E.
submitted to the EMBL Data Library, January 1994
A;Reference number: S42069
A;Accession: S42069
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-236 <GUE>
A;Cross-references: EMBL:X75855; NID:g456206; PIDN:CAA53470.1; PID:g456207
R;Walter, L.; Dirks, B.; Rothermel, E.; Heyens, M.; Szpiter, C.; Levan, G.; Gunther,
Mamm. Genome 5, 216-221, 1994
A;Title: A novel, conserved gene of the rat that is developmentally regulated in the
A;Reference number: I57015; MUID:94281747; PMID:8012111
A;Accession: I57015
A;Status: preliminary; translated from GB/EMBL/DDBU
A;Molecule type: mRNA
A;Residues: 1-236 <RES>

A;Cross-references: EMBL:X75855; NID:g456206; PIDN:CAA53470.1; PID:g456207
 A;Accession: I76675
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 2-236 <RE2>
 A;Cross-references: EMBL:X75856; NID:g456208; PIDN:CAA53471.1; PID:g456209
 C;Genetics:
 A;Gene: test
 C;Superfamily: human testicular protein TEGT

Query Match 81.4%; Score 35; DB 2; Length 236;
 Best Local Similarity 87.5%; Pred. No. 13;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 YVSSLGNL 8
 :|||||
 Db 154 FVSSLGNL 161

RESULT 3
 S28086
 gene A protein - yeast (Kluyveromyces marxianus var. drosophilum) plasmid pKD1
 C;Species: Kluyveromyces marxianus var. drosophilum
 C;Date: 07-May-1993 #sequence_revision 07-May-1993 #text_change 29-Oct-1999
 C;Accession: S28086
 R;Chen, X.J.; Sallola, M.; Falcone, C.; Bianchi, M.M.; Fukuhara, H.
 Nucleic Acids Res. 14, 4471-4481, 1986
 A;Title: Sequence organization of the circular plasmid pKD1 from the yeast Kluyveromyces
 A;Reference number: S28086; MUID:86232585; PMID:3520486
 A;Accession: S28086
 A;Status: translation not shown
 A;Molecule type: DNA
 A;Residues: 1-447 <CHE>
 A;Cross-references: EMBL:X33961; NID:g2800; PIDN:CAA27591.1; PID:g2801
 C;Genetics:
 A;Gene: A
 A;Genome: plasmid
 C;Superfamily: site-specific recombinase FLP

Query Match 81.4%; Score 35; DB 2; Length 447;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YVSSLGN 7
 :|||||
 Db 314 YVSSLGN 320

RESULT 4
 F95074
 serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR4)
 C;Species: Streptococcus pneumoniae
 C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
 C;Accession: F95074
 R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
 on, J.D.; Unayam, E.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,
 nson, T.; Hickey, E.K.; Holt, I.E.
 Science 293, 498-506, 2001
 A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
 A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
 A;Reference number: A95000; MUID:21357209; PMID:11463916
 A;Accession: F95074
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-2140 <KUR>
 A;Cross-references: GB:AE005672; PIDN:AAK74791.1; PID:g14972117; GSPDB:GN00164; TIGR:SP4
 A;Experimental source: strain TIGR4
 C;Genetics:
 A;Gene: SP0641

Query Match 81.4%; Score 35; DB 2; Length 2140;
 Best Local Similarity 77.8%; Pred. No. 1.5e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YVSSLGNLI 9
 :|||||
 Db 1555 YVSSYGNI 1563

RESULT 5
 A97942
 metalloproteinase (EC 3.4.21.-) A [imported] - Streptococcus pneumoniae (strain R6)
 C;Species: Streptococcus pneumoniae
 C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
 C;Accession: A97942
 R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.
 e, R.; LeBlanc, D.J.; Lee, D.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.
 Y. P.; Sun, P.M.; Winkler, M.E.
 J. Bacteriol. 183, 5709-5717, 2001
 A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
 A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
 A;Reference number: A97872; MUID:21429245; PMID:11544234
 A;Accession: A97942
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-2144 <KUR>
 A;Cross-references: GB:AE007317; PIDN:AAK99365.1; PID:g15458138; GSPDB:GN00174
 C;Genetics:
 A;Gene: PrtA
 C;Keywords: hydrolase; serine proteinase

Query Match 81.4%; Score 35; DB 2; Length 2144;
 Best Local Similarity 77.8%; Pred. No. 1.5e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YVSSLGNLI 9
 :|||||
 Db 1559 YVSSYGNI 1567

RESULT 6
 C72420
 hypothetical protein TM0089 - Thermotoga maritima (strain MSB8)
 C;Species: Thermotoga maritima
 C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
 C;Accession: C72420
 R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Swinn, M.L.; Dodson, R.J.; Haft, D.H.; Hi
 Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
 C.M.
 Nature 399, 323-329, 1999
 A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
 A;Reference number: A72200; MUID:99287316; PMID:10360571
 A;Accession: C72420
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-153 <ARN>
 A;Cross-references: GB:AE001695; GB:AE000512; NID:g4980569; PIDN:AAK35183.1; PID:g49
 A;Experimental source: strain MSB8
 C;Genetics:
 A;Gene: TM0089
 C;Superfamily: Thermotoga maritima hypothetical protein TM0089

Query Match 76.7%; Score 33; DB 2; Length 153;
 Best Local Similarity 77.8%; Pred. No. 21;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YVSSLGNLI 9
 :|||||
 Db 133 YVSSGLVIV 141

RESULT 7
 T49580
 probable oligomycin sensitivity conferring protein (ATP5) [imported] - Neurospora cr
 N;Alternate names: protein B208.280
 C;Species: Neurospora crassa

C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
 C/Accession: T49580
 R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura, submitted to the Protein Sequence Database, May 2000
 A/Reference number: Z25022
 A/Accession: T49580
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-220 <SCH>
 A/Cross-references: EMBL:AL355930; GSPDB:GN00116; NCSP:B208.280
 A/Experimental source: BAC clone B208; strain OR74A
 C/Genetics:
 A/Gene: NCSP:B208.280
 A/Map position: 6
 A/Introns: 48/3; 152/3; 183/3

Query Match 76.7%; Score 33; DB 2; Length 220;
 Best Local Similarity 75.0%; Pred. No. 31;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VSSLSGNLI 3
 Db 65 ISSLSGNLI 72

RESULT 8
 T11154
 Cytochrome-c oxidase (EC 1.9.3.1) chain I - hardbacked tick (*Rhipicephalus sanguineus*)
 C/Species: mitochondrion *Rhipicephalus sanguineus*
 C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jan-2000
 C/Accession: T11154
 R;Black IV, W.C.; Roehrdanz, R.L.
 Vol. Biol. Evol. 15, 1772-1785, 1998
 A/Title: Mitochondrial gene order is not conserved in arthropods: prostriate and metastriate
 A/Reference number: Z17252; MUID:99083443; PMID:9866211
 A/Accession: T11154
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-512 <BLA>
 A/Cross-references: EMBL:AF081829; NID:G4164556; PID:G4164558; PIDN:AAD05518.1
 C/Genetics:
 A/Gene: mitochondrion
 C/Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
 F/10-256/Domain: cytochrome-c oxidase chain I homology <COL>
 F/60,377/Binding site: heme a iron (His) (axial ligands) #status predicted
 F/239,289,290/Binding site: copper (His) #status predicted
 F/239-243/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
 F/243/Binding site: oxygen (Tyr) #status predicted
 F/367/Binding site: magnesium (His) (shared with chain II) #status predicted
 F/375/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 76.7%; Score 33; DB 2; Length 512;
 Best Local Similarity 66.7%; Pred. No. 80;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YVSSLSGNLI 9
 Db 451 FISSLSGLI 459

RESULT 9
 S58740
 Cytochrome-c oxidase (EC 1.9.3.1) chain I - yeast (*Hansenula wingei*) mitochondrion
 C/Species: mitochondrion *Hansenula wingei*
 C/Date: 19-Mar-1997 #sequence_revision 09-May-1997 #text_change 07-Dec-1999
 C/Accession: S58740
 R;Sekito, T.; Okamoto, K.; Kitano, H.; Yoshida, K.
 Curr. Genet. 28, 39-53, 1995
 A/Title: The complete mitochondrial DNA sequence of *Hansenula wingei* reveals new characters
 A/Reference number: S58740; MUID:96022424; PMID:8536312
 A/Accession: S58740
 A/Molecule type: DNA

A/Residues: 1-535 <SEK>
 A/Cross-references: EMBL:D31785
 A/Note: The authors translated the codon CAA for residue 54 as Gly
 C/Genetics:
 A/Gene: cox1
 A/Gene: mitochondrion
 A/Genetic code: SGC3
 A/Introns: 241/3
 C/Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
 C/Keywords: chromoprotein; copper; electron transfer; heme; iron; lipoprotein; magnesium; respiratory chain; transmembrane protein
 F/11-458/Domain: cytochrome-c oxidase chain I homology <COL>
 F/63,379/Binding site: heme a iron (His) (axial ligands) #status predicted
 F/242,291,292/Binding site: copper (His) #status predicted
 F/242-246/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
 F/246/Binding site: oxygen (Tyr) #status predicted
 F/320/Binding site: myristate (Lys) (covalent) #status predicted
 F/369/Binding site: magnesium (His) (shared with chain II) #status predicted
 F/377/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 76.7%; Score 33; DB 2; Length 535;
 Best Local Similarity 66.7%; Pred. No. 84;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YVSSLSGNLI 9
 Db 453 YVSSLSGVI 461

RESULT 10
 S46404
 vitellogenin - yellow fever mosquito
 C/Species: *Aedes aegypti* (yellow fever mosquito)
 C/Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 21-Jul-2000
 C/Accession: S46404
 R;Chat, J.S.; Cho, W.L.; Raikhel, A.S.
 J. Mol. Biol. 237, 641-647, 1994
 A/Title: Analysis of mosquito vitellogenin cDNA. Similarity with vertebrate phosphovitin
 A/Reference number: S46404; MUID:94210487; PMID:8158643
 A/Accession: S46404
 A/Molecule type: mRNA
 A/Status: preliminary
 A/Residues: 1-2139 <CHE>
 A/Cross-references: EMBL:U02548; NID:G472307; PIDN:AAA18221.1; PID:G472308

Query Match 76.7%; Score 33; DB 2; Length 2139;
 Best Local Similarity 75.0%; Pred. No. 3.9e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YVSSLSGNLI 8
 Db 801 YIRSLSGNL 808

RESULT 11
 S43056
 hypothetical protein - mouse
 C/Species: *Mus musculus* (house mouse)
 C/Date: 14-Sep-1994 #sequence_revision 26-May-1995 #text_change 05-Nov-1999
 C/Accession: S43056
 R;Sun, Y.; Hegamyer, G.; Colburn, N.H.
 Cancer Res. 54, 1139-1144, 1994
 A/Title: Molecular cloning of five messenger RNAs differentially expressed in preneoplasia
 A/Reference number: S43053; MUID:94163596; PMID:8118794
 A/Accession: S43056
 A/Status: preliminary; translation not shown
 A/Molecule type: mRNA
 A/Residues: 1-52 <SUN>
 A/Cross-references: EMBL:Z31362; NID:G467582; PIDN:CAA83237.1; PID:G467583

Query Match 74.4%; Score 32; DB 2; Length 52;
 Best Local Similarity 100.0%; Pred. No. 10;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 SSIGNLI 9
 Db 34 SSIGNLI 40

RESULT 12

I53154
 Scleraxis - mouse
 C:Species: Mus sp. (mouse)
 C>Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999
 C:Accession: I53154
 R:Csarjesi, P.; Brown, D.; Ligon, K.L.; Zycns, G.E.; Copeland, N.G.; Gilbert, D.J.; Jenk
 Development 121, 1099-1110, 1995
 A>Title: Scleraxis: a basic helix-loop-helix protein that prefigures skeletal formation
 A:Reference number: I53154; MUID:95262555; PMID:7743923
 A:Accession: I53154
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-207 <RES>
 A:Cross-references: GB:S78079; NID:g9988393; PIDN:ARB34266.1; PID:g998839

Query Match 74.4%; Score 32; DB 2; Length 207;

Best Local Similarity 55.6%; Pred. No. 48;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 YVSSIGNLI 9
 Db 126 YISLGNVL 134

RESULT 13

C64347
 hypothetical protein MJ0379 - Methanococcus jannaschii
 C:Species: Methanococcus jannaschii
 C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
 C:Accession: C64347
 R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, A.
 Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
 rsen, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
 Science 273, 1058-1073, 1996
 A:Authors: Kaine, B.P.; Bordovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C
 A>Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
 A:Reference number: A64300; MUID:96337999; PMID:8688087
 A:Accession: C64347
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-208 <HUL>
 A:Cross-references: GB:U67490; GB:L77117; NID:g2826276; PIDN:AB98369.1; PID:g1591084; T
 C:Genetics:
 A:Map position: REV345512-344886
 A:Start codon: TTG

Query Match 74.4%; Score 32; DB 2; Length 208;

Best Local Similarity 66.7%; Pred. No. 48;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 YVSSIGNLI 9
 Db 128 YVSLGDIL 136

RESULT 14

AD0513
 bis(5'-nucleosyl)-tetraphosphatase (symmetrical) (EC 3.6.1.41) STY0103 [similarity] - Sa
 C:Species: Salmonella enterica subsp. enterica serovar Typhi
 A>Note: this species has also been called Salmonella typhi
 C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
 C:Accession: AD0513
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
 th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
 , S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
 A>Title: Complete genome sequence of a multiple drug resistant Salmonella enterica s
 A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AD0513
 A>Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-282 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD01244.1; PID:g16501373; GSPDB:GN00176
 C:Genetics:

A:Gene: STY0103
 C:Superfamily: bis(5'-nucleosyl)-tetraphosphatase (symmetrical); phosphoesterase cor
 C:Keywords: hydrolase

Query Match 74.4%; Score 32; DB 2; Length 282;

Best Local Similarity 66.7%; Pred. No. 67;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 YVSSIGNLI 9
 Db 51 YVSLGNSV 59

RESULT 15

A40708
 basic-helix-loop-helix protein ato - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C>Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
 C:Accession: A40708
 R:Jarman, A.P.; Grau, Y.; Jan, L.Y.; Jan, Y.N.
 Cell 73, 1307-1321, 1993
 A>Title: atonal is a proneural gene that directs chordotonal organ formation in the
 A:Reference number: A40708; MUID:93313961; PMID:8324823
 A:Accession: A40708

A>Status: preliminary
 A:Molecule type: nucleic acid

A:Residues: 1-312 <JAR>
 A:Cross-references: GB:L36646; NID:g551565; PIDN:AAA21879.1; PID:g551566
 A>Note: sequence extracted from NCBI backbone (NCBIN:135094, NCBIP:135095)

C:Genetics:

A:Gene: FlyBase:ato
 A:Cross-references: FlyBase:FBgn0010433

Query Match 74.4%; Score 32; DB 2; Length 312;

Best Local Similarity 55.6%; Pred. No. 75;
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 YVSSIGNLI 9
 Db 303 YISALGDIL 311

Search completed: January 20, 2004, 17:03:09

Job time : 20.3435 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:41:41 ; Search time 4.74046 Seconds
(without alignments)
89.283 Million cell updates/sec

Title: US-10-089-549-1
Perfect score: 43
Sequence: 1 YVSSLGNLI 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	81.4	236	1 BIL_RAT	P50622 rattus norv
2	35	81.4	447	1 FLP_KLULA	P13783 kluyveromyc
3	33	76.7	220	1 ATP0_NEUCR	O96022 neurospora
4	33	76.7	512	1 COX1_RHISA	O98818 rhinipicaph
5	33	76.7	535	1 COX1_HANWI	P48868 hansenula w
6	33	76.7	570	1 ETPD_ACICA	P94132 acinetobact
7	33	76.7	2148	1 VIT1_AEDAE	Q16927 aedes aegypt
8	32	74.4	187	1 SCX_CHICK	P59101 gallus gall
9	32	74.4	207	1 SCX_MOUSE	O64124 mus musculu
10	32	74.4	208	1 Y379_METUA	Q57824 methanococc
11	32	74.4	282	1 APAH_SALTY	Q56018 salmonella
12	32	74.4	312	1 ATO_DROME	P48987 drosophila
13	32	74.4	348	1 MOZL_ARATH	Q92477 arabidopsis
14	32	74.4	419	1 GP63_HUMAN	Q9bzj6 homo sapien
15	32	74.4	425	1 GP63_MOUSE	O9eqq3 mus musculu
16	32	74.4	579	1 UL25_ZYVD	Q92877 varicella-z
17	32	74.4	898	1 CLC2_HUMAN	P51788 homo sapien
18	32	74.4	898	1 CLC2_RABIT	P51789 oryctolagus
19	32	74.4	902	1 CLC2_CAVPO	Q9wu45 cavia porce
20	32	74.4	907	1 CLC2_RAT	P35525 rattus norv
21	32	74.4	908	1 CLC2_MOUSE	Q9r0al mus musculu
22	32	74.4	1073	1 YAS5_SCHPO	O10141 schizosacch
23	32	74.4	1178	1 CIAC_BACTK	P05068 bacillus th
24	31	72.1	192	1 SODF_ECOLI	P09157 escherichia
25	31	72.1	192	1 SODF_SALTY	P40726 salmonella
26	31	72.1	213	1 SODF_HELP	O9zke6 helicobacte
27	31	72.1	213	1 SODF_HELPV	P43312 helicobacte
28	31	72.1	220	1 SODF_CAMCO	P53639 campylobact
29	31	72.1	220	1 SODF_CAMCE	P53640 campylobact
30	31	72.1	359	1 AROC_CHLEN	Q9z6m2 chlamydia p
31	31	72.1	415	1 Y207_METUA	Q60269 methanococc
32	31	72.1	450	1 SYSC_SCHPO	O14018 schizosacch
33	31	72.1	534	1 COX1_KLULA	P20386 kluyveromyc

34	31	72.1	575	1 ITRL_SCHPO	Q10286 schizosacch
35	31	72.1	781	1 PARC_MYCGE	P47446 mycoplasma
36	31	72.1	789	1 PARC_MYCPN	P75352 mycoplasma
37	31	72.1	805	1 CICH_TORMA	P21564 torpedo mar
38	31	72.1	809	1 CICH_TORCA	P35522 torpedo cal
39	31	72.1	1517	1 GLTB_ECOLI	P09831 escherichia
40	31	72.1	1716	1 RPAL_RAT	O54889 rattus norv
41	30	69.8	145	1 HBB1_XENBO	P07432 xenopus bor
42	30	69.8	145	1 HBB1_XENLA	P02132 xenopus lae
43	30	69.8	145	1 HBB2_XENBO	P07433 xenopus bor
44	30	69.8	146	1 HBB1_XENTR	P07429 xenopus tro
45	30	69.8	146	1 HBB2_XENTR	P08423 xenopus tro

ALIGNMENTS

RESULT 1

BIL_RAT	STANDARD;	PRT;	236 AA.
AC	P50622; Q64712;		
DT	01-OCT-1996 (Rel. 34, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Bax inhibitor-1 (BI-1) (Testis enhanced gene transcript).		
GN	TEGT OR BIL.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_TaxID=10116;		
[1]	SEQUENCE FROM N.A.		
RP	STRAIN-Sprague-Dawley; TISSUE=Testis;		
RC	MEDLINE=94281747; PubMed=8012111;		
RX	Walter L., Dirks B., Rothermel E., Heyens M., Szpirer C., Levan G.,		
RA	Guenther E.;		
RT	*A novel, conserved gene of the rat that is developmentally regulated		
RT	in the testis.*;		
RL	Mamm. Genome 5:216-221(1994).		
CC	FUNCTION: SUPPRESSOR OF APOPTOSIS (BY SIMILARITY).		
CC	SUBUNIT: INTERACTS WITH BLC2 AND BCL-XL (BY SIMILARITY).		
CC	SUBCELLULAR LOCATION: Integral membrane protein (Potential).		
CC	TISSUE SPECIFICITY: HIGHLY ABUNDANT IN TESTIS.		
CC	SIMILARITY: BELONGS TO THE BIL FAMILY.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
DR	EMBL; X75855; CAA53470.1; -		
DR	EMBL; X75856; CAA53471.1; -		
DR	PIR; S42069; S42069.		
DR	InterPro; IPR006213; Bax inhbt1.		
DR	InterPro; IPR006214; UPF0005.		
DR	Pfam; PF01027; UPF0005; 1.		
DR	PROSITE; PS01243; BIL; 1.		
KW	Apoptosis; Transmembrane.		
FT	TRANSMEM 30 50		POTENTIAL.
FT	TRANSMEM 53 73		POTENTIAL.
FT	TRANSMEM 88 108		POTENTIAL.
FT	TRANSMEM 112 132		POTENTIAL.
FT	TRANSMEM 139 159		POTENTIAL.
FT	TRANSMEM 166 186		POTENTIAL.
FT	TRANSMEM 206 226		POTENTIAL.
SQ	SEQUENCE 236 AA; 26358 MW; FPA412ECIDC87537 CRC64;		

Query Match 81.4%; Score 35; DB 1; Length 236;
Best Local Similarity 87.5%; Pred. No. 7.3;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;


```

OX RN NCBI_TaxID=5141;
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A / FGSC 987;
RX PubMed=12655011;
RA Maehaupt G., Montrone C., Haase D., Mewes H.-M., Aign V.,
RA Honeisel J.D., Fartmann B., Nyakatura G., Kempken F., Maier J.,
RA Schulte U.;
RT "What's in the genome of a filamentous fungus? Analysis of the
RT Neurospora genome sequence.";
RL Nucleic Acids Res. 31:1944-1954(2003).
CC -!- FUNCTION: This protein seems to be part of the stalk that links
CC CF(0) to CF(1). It either transmits conformational changes from
CC CF(0) into CF(1) or is implicated in proton conduction (by
CC similarity).
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) [in] = ADP + phosphate +
CC H(+) (out).
CC -!- SUBUNIT: F-type ATPases have 2 components, CF(1) - the catalytic
CC core - and CF(0) - the membrane proton channel. CF(1) has five
CC subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1). CF(0)
CC has three main subunits: A, B and C (by similarity).
CC -!- SUBCELLULAR LOCATION: Mitochondrial.
CC -!- SIMILARITY: BELONGS TO THE ATPASE DELTA CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL355930; CAB91368.2; -
DR F1R; T49580; T49580.
DR InterPro; IPR000711; ATPsynth_OSCP.
DR Pfam; PF00213; OSCP; 1.
DR PRINTS; PR00125; ATPASEDELTA.
DR TIGRFAM; TIGR01145; ATP synt delta; 1.
DR PROSITE; PS00389; ATPASE_DELTA; 1.
KW Hydrolyase; ATP synthetase; CF(1); Hydrogen ion transport;
KW Mitochondrion; Transit peptide.
FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL).
FT CHAIN ? 220 ATP SYNTHASE OLIGOMYCIN SENSITIVITY
FT CONFERRAL PROTEIN.
FT SEQUENCE 220 AA; 23047 MW; 7C6E82BD511C2C81 CRC64;
Query Match 76.7%; Score 33; DB 1; Length 220;
Best Local Similarity 75.0%; Pred. No. 17;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 2 VSSLSGNLI 9
Db :|||||:
65 ISSLGNLL 72
RESULT 4
COX1_RHISA
ID COX1_RHISA STANDARD; PRT; 512 AA.
AC G99818;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1).
GN COI.
OS Rhinipcephalus sanguineus (Brown dog tick).
OG Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Parasitiformes; Ixodida; Ixodidae; Rhinipcephalus.
OX NCBI_TaxID=34632;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99083443; PubMed=9866211;
RA Black W.C. IV, Roehrdanz R.L.;

```

```

QY 1 YVSSLSGNL 8
Db :|||||:
154 YVSSLSGNL 161
RESULT 2
FLP_KLUJLA
ID FLP_KLUJLA STANDARD; PRT; 447 AA.
AC P13783;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Recombinase FLP protein.
GN A.
OS Kluyveromyces lactis (Yeast).
OG Plasmid pKDL.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=28985;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86232585; PubMed=3520486;
RA Chen X.J., Saliola M., Falcone C., Bianchi M.M., Fukuhara H.;
RT "Sequence organization of the circular plasmid pKDL from the yeast
RT Kluyveromyces drosophilaram.";
RL Nucleic Acids Res. 14:4471-4481(1986).
CC -!- FUNCTION: FLP catalyzes recombination between the large inverted
CC repetitions of the plasmid.
CC -!- SIMILARITY: BELONGS TO THE "PHAGE" INTEGRASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X03961; CAAT27591.1; -
DR F1R; S28086; S28086.
DR InterPro; IPR005626; FLP.
DR Pfam; PF05202; FLP_C; 1.
DR Pfam; PF03930; FLP_N; 1.
DR ProDom; PD010043; FLP; 1.
KW DNA recombination; DNA integration; Plasmid.
FT ACT_SITE 336 TRANSIENT COVALENT LINKAGE TO DNA DURING
STRAND CLEAVAGE AND RETOINING (BY
SIMILARITY).
FT ACT_SITE 336
FT SEQUENCE 447 AA; 51110 MW; 89D08AB52718A104 CRC64;
Query Match 81.4%; Score 35; DB 1; Length 447;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 YVSSLSGN 7
Db :|||||:
314 YVSSLSGN 320
RESULT 3
ATPO_NEUCR
ID ATPO_NEUCR STANDARD; PRT; 220 AA.
AC Q9P602;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE ATP synthase oligomycin sensitivity conferral protein, mitochondrial
DE precursor (EC 3.6.3.14) (OSCP) (ATP synthase chain 5).
GN ATP-5 OR B208.150.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.

```

RT Mitochondrial gene order is not conserved in arthropods: prostrate
 RT and metazoa tick mitochondrial genomes.";
 RL Mol. Biol. Evol. 15:1772-1785 (1998).
 CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
 CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
 CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
 CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
 CC AND COPPER B.
 CC -!- CATALYTIC ACTIVITY: 4 ferrocytochrome c + O(2) = 4 ferrocytochrome
 CC c + 2 H(2)O.
 CC -!- PATHWAY: Respiratory chain; terminal step.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
 CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
 CC -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; AF081829; AAD05518.1; -
 DR PIR; T11154; T11154.
 DR HSSP; P00396; 20CC.
 DR InterPro; IPR000883; COX1.
 DR Pfam; PF00115; COX1; 1.
 DR PRINTS; PR01165; CYCOXIDASE1.
 DR PROSITE; PS00077; COX1; 1.
 KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
 KW Respiratory chain; Inner membrane.
 FT METAL 60 60 IRON (HEME A AXIAL LIGAND) (PROBABLE).
 FT METAL 239 239 COPPER B (PROBABLE).
 FT METAL 243 243 COPPER B (PROBABLE).
 FT METAL 289 289 COPPER B (PROBABLE).
 FT METAL 290 290 COPPER B (PROBABLE).
 FT METAL 375 375 IRON (HEME A3 AXIAL LIGAND) (PROBABLE).
 FT METAL 377 377 IRON (HEME A AXIAL LIGAND) (PROBABLE).
 FT CROSSLINK 239 243 1'-histidyl-3'-tyrosine (By similarity).
 SQ SEQUENCE 512 AA; 57125 MW; 5306374C4DF11AB6 CRC64;
 Query Match 76.7%; Score 33; DB 1; Length 512;
 Best Local Similarity 66.7%; Pred. No. 42;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 YVSSLGSLI 9
 DB 451 FISSLSGLI 459
 RESULT 5
 COX1_HANWI STANDARD; PRT; 535 AA.
 AC P48668;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1).
 GN COX1.
 OS Hansenula wingei (Yeast).
 OG Mitochondrion.
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Pichia.
 OX NCBI_TaxID=4907;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=21;
 RA Sekito T., Okamoto K., Kitano H., Yoshida K.;
 RT "Yeast Hansenula wingei mitochondria genome's complete DNA sequence
 RT demonstrated unique characteristics."

RL Nucleic Acids Symp. Ser. 31:233-234 (1994).
 CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
 CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
 CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
 CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
 CC AND COPPER B.
 CC -!- CATALYTIC ACTIVITY: 4 ferrocytochrome c + O(2) = 4 ferrocytochrome
 CC c + 2 H(2)O.
 CC -!- PATHWAY: Respiratory chain; terminal step.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
 CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
 CC -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; D31785; BAA06563.2; -
 DR PIR; S58740; S58740.
 DR HSSP; P98002; IARI.
 DR InterPro; IPR000883; COX1.
 DR Pfam; PF00115; COX1; 1.
 DR PRINTS; PR01165; CYCOXIDASE1.
 DR PROSITE; PS00077; COX1; 1.
 KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
 KW Respiratory chain; Inner membrane.
 FT METAL 63 63 IRON (HEME A AXIAL LIGAND) (PROBABLE).
 FT METAL 242 242 COPPER B (PROBABLE).
 FT METAL 246 246 COPPER B (PROBABLE).
 FT METAL 291 291 COPPER B (PROBABLE).
 FT METAL 292 292 COPPER B (PROBABLE).
 FT METAL 377 377 IRON (HEME A3 AXIAL LIGAND) (PROBABLE).
 FT METAL 379 379 IRON (HEME A AXIAL LIGAND) (PROBABLE).
 FT CROSSLINK 242 246 1'-histidyl-3'-tyrosine (By similarity).
 SQ SEQUENCE 535 AA; 59025 MW; 97F7C4EFAD1AD50A CRC64;
 Query Match 76.7%; Score 33; DB 1; Length 535;
 Best Local Similarity 66.7%; Pred. No. 44;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 YVSSLGSLI 9
 DB 453 YVSSIGSLI 461
 RESULT 6
 EFTD ACICA STANDARD; PRT; 570 AA.
 ID EFTD ACICA
 AC P94132;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable electron transfer flavoprotein-ubiquinone oxidoreductase
 DE (EC 1.5.5.1) (EFT-OO) (EFT-ubiquinone oxidoreductase) (EFT
 DE dehydrogenase) [Electron-transferring-flavoprotein dehydrogenase].
 OS Acinetobacter calcoaceticus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Moraxellaceae; Acinetobacter.
 OX NCBI_TaxID=471;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BD413 / ADP1;
 RX MEDLINE=97440147; PubMed=9294455;
 RA Williams P.A., Shaw L.E.;
 RT "muck, a gene in Acinetobacter calcoaceticus ADP1 (BD413), encodes the
 RT ability to grow on exogenous cis,cis-muconate as the sole carbon
 RT source."

RL J. Bacteriol. 179:5935-5942(1997).
 CC -1- FUNCTION: ACCEPTS ELECTRONS FROM ETF AND REDUCES UBIQUINONE.
 CC -1- CATALYTIC ACTIVITY: Reduced ETF + ubiquinone = ETF + ubiquinol.
 CC -1- COFACTOR: FAD AND A 4FE-4S CLUSTER.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U87258; AAC27118.1; -
 CC InterPro: IPR003042; Rng_mnxyogenase.
 CC Pfam; PF05187; ETFD; 1
 CC PRINTS; PA00420; RINGNOXGNASE.
 CC Oxidoreductase; Electron transport; Flavoprotein; FAD; Iron-sulfur;
 KW 4FE-4S; Ubiquinone.
 FT NP BIND 13 27 FAD (ADP PART) (POTENTIAL).
 FT METAL 515 515 IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT METAL 539 539 IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT METAL 542 542 IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT METAL 545 545 IRON-SULFUR (4FE-4S) (POTENTIAL).
 SQ SEQUENCE 570 AA; 62995 MW; 0107DD88E78A7DES CRC64;
 Query Match 76.7%; Score 33; DB 1; Length 570;
 Best Local Similarity 66.7%; Pred. No. 47;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Qy i YVSSIGNLI 9
 Db 117 YVSSIGNLV 125
 RESULT 7
 ID VITL AEDAE STANDARD; PRT: 2148 AA.
 AC Q16927; Q16932;
 DT 15-JUN-1998 (Rel. 36, Created)
 DT 15-JUN-1998 (Rel. 36, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Vitellogenin A1 precursor (VG) (PVL1) [Contains: Vitellin light chain
 DE (VL); Vitellin heavy chain (VH)].
 GN VGAL.
 OS Aedes aegypti (Yellowfever mosquito).
 CC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 CC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes.
 CC NCBI_TaxID=7159;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=96035842; PubMed=7550249;
 RA Romans P., Tu Z., Ke Z., Hagedorn H.H.;
 RT Analysis of a vitellogenin gene of the mosquito, Aedes aegypti and
 RT comparisons to vitellogenins from other organisms.";
 RL Insect Biochem. Mol. Biol. 25:939-958(1995).
 RN [2]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 17-24 AND 469-477.
 RC STRAIN=UGALS; TISSUE=Fat body;
 RX MEDLINE=94210487; PubMed=8158643;
 RA Chen J.-S., Cho W.-L., Raikhel A.S.;
 RT Analysis of mosquito vitellogenin cDNA. Similarity with vertebrate
 RT phosphatins and arthropod serum proteins.";
 RL J. Mol. Biol. 237:641-647(1994).
 RN [3]
 RP TISSUE=Fat body;
 RC MEDLINE=90277688; PubMed=2351682;
 RA Dhadialla T.S., Raikhel A.S.;
 RT Biosynthesis of mosquito vitellogenin.";
 RL J. Biol. Chem. 265:9924-9933(1990).
 CC -1- FUNCTION: PRECURSOR OF THE EGG-YOLK PROTEINS THAT ARE SOURCES OF
 CC NUTRIENTS DURING EMBRYONIC DEVELOPMENT. MAY SUPPLY AROMATIC AMINO

CC ACIDS TO THE CUTICLE OF RAPIDLY DEVELOPING EMBRYOS.
 CC -1- TISSUE SPECIFICITY: PRODUCED BY THE FAT BODY, WHERE IT IS CLEAVED
 CC IN THE ROUGH ENDOPLASMIC RETICULUM OR CIS-GOLGI BEFORE BEING
 CC SECRETED INTO HEMOLYPH. IT IS THEN SEQUESTERED BY A SINGLE CLASS
 CC OF RECEPTOR MEDIATED ENDOCYTOSIS IN THE OVARY.
 CC -1- INDUCTION: SYNTHESIZED ONLY BY SEXUALLY MATURE FEMALE AFTER
 CC INGESTION OF BLOOD.
 CC -1- PTM: GLYCOSYLATED, PHOSPHORYLATED AND SULFATED. THE LARGE SUBUNIT
 CC IS SULFATED MORE EXTENSIVELY THAN THE SMALL ONE.
 CC -1- POLYMORPHISM: ALLELIC VARIATIONS DETECTED IN THE MOSQUITO
 CC POPULATION.
 CC -1- SIMILARITY: Contains 1 WVD domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L41842; AAA99486.1; -
 CC EMBL; U02548; AAA18221.1; -
 CC InterPro; IPR001747; Lipid transprt_N.
 CC InterPro; IPR001846; WVD_D
 CC Pfam; PF01347; Vitellogenin_N; 1.
 CC Pfam; PF00094; wvd; 1.
 CC SMART; SM00638; LPD_N; 1.
 CC SMART; SM00216; WVD; 1.
 CC Glycoprotein; Phosphorylation; Storage protein; Signal; Polymorphism;
 KW Sulfation.
 FT SIGNAL 1 16
 FT CHAIN 17 2148
 FT CHAIN 17 468
 FT CHAIN 469 2148
 FT MOD_RES 138 138 SULFATION (POTENTIAL).
 FT MOD_RES 142 142 SULFATION (POTENTIAL).
 FT MOD_RES 1046 1046 SULFATION (POTENTIAL).
 FT MOD_RES 1049 1049 SULFATION (POTENTIAL).
 FT MOD_RES 1053 1053 SULFATION (POTENTIAL).
 FT MOD_RES 1542 1542 SULFATION (POTENTIAL).
 FT MOD_RES 1543 1543 SULFATION (POTENTIAL).
 FT MOD_RES 1549 1549 SULFATION (POTENTIAL).
 FT MOD_RES 1716 1716 SULFATION (POTENTIAL).
 FT MOD_RES 1785 1785 SULFATION (POTENTIAL).
 FT MOD_RES 1788 1788 SULFATION (POTENTIAL).
 FT MOD_RES 1801 1801 SULFATION (POTENTIAL).
 FT MOD_RES 1803 1803 SULFATION (POTENTIAL).
 FT MOD_RES 1867 1867 SULFATION (POTENTIAL).
 FT CARBOHYD 86 86 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 104 104 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 339 339 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 370 370 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 414 414 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 493 493 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 517 517 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 566 566 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 742 742 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 760 760 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1119 1119 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1212 1212 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1315 1315 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1631 1631 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1675 1675 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1956 1956 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT DOMAIN 417 440 POLY-SER.
 FT DOMAIN 496 512 POLY-SER.
 FT DOMAIN 520 548 POLY-SER.
 FT DOMAIN 2007 2025 POLY-SER.
 FT DOMAIN 2033 2042 POLY-SER.
 FT CONFLICT 39 47 MISSING (IN REF. 2).
 FT CONFLICT 906 906 R -> S (IN REF. 2).
 FT CONFLICT 1367 1367 F -> Y (IN REF. 2).

```

FT CONFLICT 1576 1576 C -> Y (IN REF. 2).
FT CONFLICT 1757 1757 N -> Y (IN REF. 2).
FT CONFLICT 1935 1935 A -> P (IN REF. 2).
FT CONFLICT 2031 2031 G -> S (IN REF. 2).
SQ SEQUENCE 2148 AA; 250249 MW; 2AB4DFC63AC766AF CRC64;

Query Match 76.7%; Score 33; DB 1; Length 2148;
Best Local Similarity 75.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YVSSUGNL 8
   |:|:|:|:|
Db 810 YVSSUGNL 817

RESULT 8
SCX CHICK STANDARD; PRT; 187 AA.
AC P59101;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Basic helix-loop-helix transcription factor scleraxis.
GN SCX.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21469794; PubMed=11585810;
RA Schweitzer R., Chyung J.H., Murtough L.C., Brent A.E., Rosen V.,
RA Olson E.N., Lassar A., Tabin C.J.;
RT "Analysis of the tendon cell fate using Scleraxis, a specific marker
RT for tendons and ligaments."
RL Development 128:3855-3866(2001).
CC -!- FUNCTION: Plays an early essential role in mesoderm formation, as
CC well as a later role in formation of somite-derived chondrogenic
CC lineages.
CC -!- SUBUNIT: Efficient DNA binding requires dimerization with another
CC BHLH protein. Dimerizes and binds the E-box consensus sequence
CC with E12.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- TISSUE SPECIFICITY: Expressed in the intersomitic, the superficial
CC proximo-medial limb mesenchyme and the subectodermal mesenchyme.
CC -!- DEVELOPMENTAL STAGE: At stage 21, expressed in leg buds in a
CC superficial proximo-medial domain, expression is enhanced by stage
CC 23. By stage 29, expression is elaborated to include both leg and
CC wing. By stage 35, expressed in all muscle-to-bone attachment
CC sites. Expressed also in a wing aponeurosis, a tendinous element
CC arranged in flattened bands, and in some of the flattened sheets
CC of connective tissue associated with muscle.
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF505881; RAM33337.1; -
CC InterPro; IPR001092; HLH_basic.
CC Pfam; PF03010; HLH; 1.
CC SMART; SM00353; HLH; 1.
CC PROSITE; PS00038; HLH 1; 1.
CC PROSITE; PS00888; HLH 2; 1.
CC Transcription regulation; Activator; Developmental protein;
CC Nuclear protein; DNA-binding.
CC DNA_BIND 67 79

```

```

FT DOMAIN 80 120 HELIX-LOOP-HELIX MOTIF.
FT DOMAIN 145 149 POLY-GLY.
SQ SEQUENCE 187 AA; 20562 MW; F789683077C2FC1B CRC64;

Query Match 74.4%; Score 32; DB 1; Length 187;
Best Local Similarity 55.6%; Pred. No. 24;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 YVSSUGNL 9
   |:|:|:|:|
Db 115 YVSSUGNL 123

RESULT 9
SCX MOUSE STANDARD; PRT; 207 AA.
ID SCX_MOUSE
AC Q64124;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Basic helix-loop-helix transcription factor scleraxis.
GN SCX.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9526255; PubMed=7743923;
RA Cserjesi P., Brown D., Ligon K.L., Lyons G.E., Copeland N.G.,
RA Gilbert D.J., Jenkins N.A., Olson E.N.;
RT "Scleraxis: a basic helix-loop-helix protein that prefigures skeletal
RT formation during mouse embryogenesis."
RL Development 121:1099-1110(1995).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=99406627; PubMed=10477299;
RA Brown D., Wagner D., Li X.-Q., Richardson J.A., Olson E.N.;
RA "Dual role of the basic helix-loop-helix transcription factor
RT scleraxis in mesoderm formation and chondrogenesis during mouse
RT embryogenesis."
RL Development 126:4317-4329(1999).
CC -!- FUNCTION: Plays an early essential role in mesoderm formation, as
CC well as a later role in formation of somite-derived chondrogenic
CC lineages.
CC -!- SUBUNIT: Efficient DNA binding requires dimerization with another
CC BHLH protein. Dimerizes and binds the E-box consensus sequence
CC with E12.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- TISSUE SPECIFICITY: Expressed in mesenchymal precursors of
CC cartilage and in connective tissue. Highly expressed in tendons in
CC the limb, tongue and diaphragm and in cartilage of the bronchi.
CC -!- DEVELOPMENTAL STAGE: At 6.0-6.5 dpc, expressed throughout the egg
CC cylinder with highest expression in the epiblast. At 9.5-10.5 dpc,
CC expressed in the lateral sclerotome and in mesenchymal cells of
CC the limb buds and body wall. At 11.0 dpc, expressed in a metameric
CC pattern extending along the length of the embryo. By 11.5 dpc,
CC expression in the developing vertebrae and intervertebral disks is
CC extended caudally. High expression was seen in precursors of the
CC ribs and bones of the limbs. Expression in progenitors of the ribs
CC and the axial and appendicular skeleton becomes downregulated when
CC ossification is initiated.
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC

```

Wed Jan 21 11:28:18 2004

us-10-089-549-1.rap

DR EMBL; S78079; AAB34266.1; --
 DR MGD; MGI:102934; SCX.
 DR InterPro; IPR001592; HLH_basic.
 DR Pfam; PF00010; HLH_1.
 DR SMART; SM00353; HLH; 1.
 DR PROSITE; PS00038; HLH_1; 1.
 DR PROSITE; PS00888; HLH_2; 1.
 KW Nuclear protein; DNA-binding.
 FT DNA BIND 78 90 BASIC DOMAIN.
 FT DOMAIN 91 131 HELIX-LOOP-HELIX MOTIF.
 FT DOMAIN 161 170 POLY-PRO.
 SQ SEQUENCE 207 AA; 22239 MW; 8340052200A95E99 CRC64;
 Query Match 74.4%; Score 32; DB 1; Length 207;
 Best Local Similarity 55.6%; Pred. No. 26;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 YVSSLGNLI 9
 ID Y379 METJA STANDARD; PRT; 208 AA.
 AC Q57824;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MJ0379.
 GN MJ0379.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_TaxID=2193;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=868087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kervage A.R., Dougherty B.A., Tomb J.-P., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinschenk K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RT Science 273:1058-1073 (1996).
 RL Science 273:1058-1073 (1996).
 CC -!- SIMILARITY: IN THE C-TERMINAL, TO M.JANNASCHII MJ0723.

RESULT 10

Y379 METJA
 ID Y379 METJA STANDARD; PRT; 208 AA.
 AC Q57824;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MJ0379.
 GN MJ0379.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_TaxID=2193;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=868087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kervage A.R., Dougherty B.A., Tomb J.-P., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinschenk K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RT Science 273:1058-1073 (1996).
 RL Science 273:1058-1073 (1996).
 CC -!- SIMILARITY: IN THE C-TERMINAL, TO M.JANNASCHII MJ0723.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U67490; AAB98369.1; --
 DR PIR; C64347; C64347.
 DR TIGR; MJ0379; --
 DR InterPro; IPR000485; HTH_AsmC.
 DR PRINTS; PR00033; HTHASNC.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 208 AA; 24080 MW; 163D:55C47FA2623 CRC64;

Query Match 74.4%; Score 32; DB 1; Length 208;
 Best Local Similarity 66.7%; Pred. No. 27;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 YVSSLGNLI 9
 ID Y379 METJA STANDARD; PRT; 282 AA.
 AC Q56018;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Bis(5'-nucleosyl)-tetraphosphate, symmetrical (EC 3.6.1.41)
 DE (Diadenosine tetraphosphatase) (Ap4A hydrolase) (Diadenosine 5',5''-
 DE P1-P4-tetraphosphate pyrophosphohydrolase).
 GN AP4A OR STM0088 OR STY0103 OR T0091.
 OS Salmonella typhimurium, and
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=602, 601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.typhimurium; STRAIN=LT2;
 RA Smith R.L., Ahuja D., Maguire M.E.;
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.

RESULT 11

APAH SALTY
 ID AP4H SALTY STANDARD; PRT; 282 AA.
 AC Q56018;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Bis(5'-nucleosyl)-tetraphosphate, symmetrical (EC 3.6.1.41)
 DE (Diadenosine tetraphosphatase) (Ap4A hydrolase) (Diadenosine 5',5''-
 DE P1-P4-tetraphosphate pyrophosphohydrolase).
 GN AP4A OR STM0088 OR STY0103 OR T0091.
 OS Salmonella typhimurium, and
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=602, 601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.typhimurium; STRAIN=LT2;
 RA Smith R.L., Ahuja D., Maguire M.E.;
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534947; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
 LT2.";
 RT Nature 413:852-856 (2001).
 RL Nature 413:852-856 (2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.typhi; STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677609;
 RA Parhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrall B.G.;
 RA "Complete genome sequence of a multiple drug resistant Salmonella
 enterica serovar Typhi CT18.";
 RT Nature 413:848-852 (2001).
 RL Nature 413:848-852 (2001).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;
 RA "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
 and CT18.";
 RT Bacteriol. 185:2330-2337 (2003).
 RL J. Bacteriol. 185:2330-2337 (2003).
 CC -!- FUNCTION: Hydrolyzes diadenosine 5',5''-P1,P4-tetraphosphate to
 CC yield ADP (By similarity).
 CC -!- CATALYTIC ACTIVITY: P(1),P(4)-bis(5'-adenosyl)tetraphosphate +
 CC H2O = 2 ADP.
 CC -!- SIMILARITY: Belongs to the Ap4A hydrolase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W., Foeller C., Gabriellian A.E., Garg N.S.C., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclet J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirskaas R., Tector C., Turner K., Turner E., Wang A.H., Wang X., Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

"The genome sequence of *Drosophila melanogaster*."

Science 287:2185-2195(2000).

[3]

FUNCTION.

MEDLINE=94255014; PubMed=8196767;

Jarman A.P., Grell E.H., Ackerman L., Jan L.Y., Jan Y.N.;

"Atrial is the proneural gene for Drosophila photoreceptors.";

Nature 369:398-400(1994).

-!- FUNCTION: DEVELOPMENTAL PROTEIN INVOLVED IN NEUROGENESIS. REQUIRED FOR THE FORMATION OF CHORDONAL ORGANS AND PHOTORECEPTORS. SEEMS TO BIND TO E BOXES. SPECIFICALLY REQUIRED FOR THE PHOTORECEPTOR R8 SELECTION.

-!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER BHLH PROTEIN. FORMS HETERODIMER WITH DAUGHTERLESS.

-!- SUBCELLULAR LOCATION: Nuclear.

-!- TISSUE SPECIFICITY: PRONEURAL CLUSTERS AND SENSE ORGAN PRECURSORS OF THE CHORDONAL ORGANS. OPTIC FURROW OF THE EYE-ANTENNAL DISK AND DEVELOPING BRAIN LOBE.

-!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF TRANSCRIPTION FACTORS. "ATONAL" SUBFAMILY.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announcement_xr_send_an_email_to_license@sib-sib.ch).

entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U24176; AAY79340.1; -
EMBL; A5008697; AAL19C52.1; -
EMBL; AL627265; CAD01244.1; -
EMBL; A5016834; AAO67824.1; -
DR StyGene; SG10537; apah.
DR HAMAP; MF 00193; -; 1.
DR InterPro; IPR004617; Apah.
DR InterPro; IPR004843; M-prestrase
DR Pfam; PF00149; Metallophos; 1.
DR ProDom; PD00252; T_ptase_apah; 1.
DR TIGRFAMS; TIGR00668; apah; 1.
KW Hydrolase; Complete proteome.
FT CONFLICT 110 110 V -> I (IN REF. 1).
FT CONFLICT 168 168 S -> R (IN REF. 1).
FT CONFLICT 171 171 A -> G (IN REF. 1).
FT CONFLICT 185 185 Y -> F (IN REF. 1).
FT CONFLICT 198 198 A -> E (IN REF. 1).
FT CONFLICT 202 202 N -> E (IN REF. 1).
FT CONFLICT 218 220 SEA -> AEE (IN REF. 1).
FT CONFLICT 252 252 E -> T (IN REF. 1).
FT CONFLICT 271 282 QMDMGGEAVNA -> HKDGEAAAS (IN REF. 1).
SQ SEQUENCE 282 AA; 31430 MW; B5B701E39CA7869E CRC64;

Query Match 74.4%; Score 32; DB 1; Length 282;
Best Local Similarity 66.7%; Pred. No. 36;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 YVSSLGNLI 9
|||
51 YVKSLSNV 59

RESULT 12
ID ATO DROME STANDARD; PRT; 312 AA.
AC P48987; Q9VHU0.
DT 01-FEB-1996 (Rel. 33, Created)
DI 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Atonal protein.
GN ATO OR CG7508.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
ON NCBI_TaxID=7227;
OX [1]
RP SEQUENCES FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC STRAIN=Oregon-R;
RX MEDLINE=93313961; PubMed=8324823;
RA Jarman A.P., Grau Y., Jan L.Y., Jan Y.N.;
RT "Atonal is a proneural gene that directs chordotonal organ formation
in the Drosophila peripheral nervous system.";
RL Cell 73:1307-1321 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
AN Anantides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Borkan M.R., Bouck J., Brockstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
FT DOMAIN 268 308 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).

Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabriellian A.B., Gaig N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacile J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner E., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
RN [3]
RP FUNCTION.
RX MEDLINE=94255014; PubMed=8196767;
RA Jarman A.P., Grell E.H., Ackerman L., Jan L.Y., Jan Y.N.;
RT "Atonal is the proneural gene for Drosophila photoreceptors.";
RL Nature 369:398-400 (1994).
CC -!- FUNCTION: DEVELOPMENTAL PROTEIN INVOLVED IN NEUROGENESIS. REQUIRED
CC FOR THE FORMATION OF CHORDOTONAL ORGANS AND PHOTORECEPTORS. SEEMS
CC TO BIND TO E BOXES. SPECIFICALLY REQUIRED FOR THE PHOTORECEPTOR R8
CC SELECTION.
CC -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN. FORMS HETERODIMER WITH DAUGHTERLESS.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: PRONEURAL CLUSTERS AND SENSE ORGAN PRECURSORS
CC OF THE CHORDOTONAL ORGANS, OPTIC FURROW OF THE EYE-ANTENNAL DISK
CC AND DEVELOPING BRAIN LOBE.
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. "ATONAL" SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; L36646; AAA21879.1; -
EMBL; A3003678; AAF54209.1; -
DR PIR; A40708; A40708.
DR FlyBase; FBgn0010433; ato.
DR GO; GO:0007420; P:brain development; IMP.
DR GO; GO:0007173; P:EGF receptor signaling pathway; IGI.
DR GO; GO:0007605; P:hearing; IMP.
DR GO; GO:0007422; P:peripheral nervous system development; NAS.
DR GO; GO:0045464; P:R8 cell fate specification; NAS.
DR GO; GO:0007224; P:smoothed receptor signaling pathway; IGI.
DR InterPro; IPR001092; HLH_basic.
DR Pfam; PF00010; HLH; 1.
DR SMART; SM00353; HLH; 1.
DR PROSITE; PS00038; HLH 1; FALSE_NEG.
DR PROSITE; PS00888; HLH 2; 1.
KW Neurogenesis; Differentiation; Developmental protein; Nuclear protein;
Transcription regulation; DNA-binding.
FT DNA BIND 255 267 BASIC DOMAIN.
FT DOMAIN 268 308 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).

FT CONFLICT 149 149 G -> A (IN REF. 1).
SQ SEQUENCE 312 AA; 34-16 MW; 069479287438F456 CRC64;

Query Match 74.4%; Score 32; DB 1; Length 312;
Best Local Similarity 55.6%; Pred. No. 40;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 YVSSIGNLI 9
|:|:|:|:
Db 303 YISALGDLI 311

RESULT 13
MO2L ARATH STANDARD; PRT; 348 AA.
AC O92Q77;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical MO25-like protein At2G03410.
GN AT2G03410 OR T4X8.16.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
CX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA SPRAIN-cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
Fuji C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
Buell C.R., Ketchum K.A., Lee J.J., Rensing C.M., Koo H.L.,
Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Cressy T.H.,
Goodman H.M., Somerville C.R., Copenhaver G.P., Press D.,
Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
Venter J.C.;
RA "Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana.";
RT Nature 402:761-768(1999).
RL Nature 402:761-768(1999).
CC -!- SIMILARITY: Belongs to the Mo25 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AC006284; AAD17435.1; -;
DR PIR; B84448; B84448.
DR InterPro; IPR004892; Mo25.
DR Pfam; PF03204; Mo25; 1.
KW Hypothetical protein.
SQ SEQUENCE 348 AA; 40000 MW; AB1D92EA2E2B900E CRC64;

Query Match 74.4%; Score 32; DB 1; Length 348;
Best Local Similarity 87.5%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY - YVSSIGNLI 8
|:|:|:|:
Db 255 YVSSLDNL 262

RESULT 14
GP63 HUMAN STANDARD; PRT; 419 AA.
ID GP63 HUMAN
AC Q9B2J6; Q9UJH3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT DE 28-FEB-2003 (Rel. 41, Last annotation update)
GN Probable G protein-coupled receptor GPR63 (PSP24-beta) (PSP24-2).
OS GPR63 OR PSP24B.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21105913; PubMed=11165367;
RA Lee D.K., George S.R., Cheng R., Nguyen T., Liu Y., Brown M.,
Lynch K.R., O'Dowd B.F.;
RT "Identification of four novel human G protein-coupled receptors
expressed in the brain";
RL Brain Res. Mol. Brain Res. 86:13-22(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20483613; PubMed=11027574;
RA Kawasawa Y., Kume K., Nakade S., Haga H., Izumi T., Shimizu T.;
RT "Brain-specific expression of novel G-protein-coupled receptors, with
homologies to Xenopus PSP24 and human GPR45.";
RL Biochem. Biophys. Res. Commun. 276:952-956(2000).
RN [3]
RP SEQUENCE FROM N.A.
RA Williams S.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: Orphan receptor. May play a role in brain function.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: Expressed in brain; detected in the frontal
cortex, with lower levels in the thalamus, caudate, hypothalamus
and midbrain.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AF317654; AAK12639.2; -;
DR EMBL; AB030566; BAB20031.1; -;
DR EMBL; AL033379; CAB55871.1; -;
DR Genew; HGNC:13302; GPR63.
DR MIM; 606915; -;
DR InterPro; IPR000276; GPCR_Rhodopsn.
DR Pfam; PF00001; 7tm 1; 1.
DR PRINTS; PR00237; GPCRHHODOPSN.
DR PROSITE; PS00237; G-PROTEIN RECEPTOR; 1.
DR PROSITE; PS00262; G-PROTEIN RECEPTOR; 2.
KW G-protein coupled receptor; Transmembrane; Glycoprotein.
FT DOMAIN 1 81
FT EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 82 104
FT DOMAIN 105 115
FT TRANSMEM 116 138
FT TRANSMEM 139 157
FT DOMAIN 158 177
FT TRANSMEM 178 196
FT DOMAIN 197 216
FT TRANSMEM 217 240
FT DOMAIN 241 263
FT TRANSMEM 264 315
FT TRANSMEM 316 338
FT DOMAIN 339 352
FT TRANSMEM 353 375
FT DOMAIN 376 419
FT CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 16 16
FT CARBOHYD 28 28
FT CARBOHYD 62 62
FT SEQUENCE 419 AA; 47577 MW; 17418CD0AE1D6C8D CRC64;

Query Match 74.4%; Score 32; DB 1; Length 419;
Best Local Similarity 66.7%; Pred. No. 54;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 YVSSIGNLI 9
DB 93 FVSFLGNLV 101

QY 1 YVSSIGNLI 9
DB 99 FVSFLGNLV 107

Search completed: January 20, 2004, 16:54:03
Job time : 9.74046 secs

```
RESULT 15
GP63_MOUSE
ID GP63_MOUSE STANDARD; PRT; 425 AA.
AC Q9BQQ3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable G protein-coupled receptor GPR63 (PSP24-beta) (PSP24-2).
GN GPR63.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20483613; PubMed=11027574;
RA Kawasaki Y., Kume K., Nakade S., Haga H., Izumi T., Shimizu T.;
RT "Brain-specific expression of novel G-protein-coupled receptors, with
RL homologues to Xenopus PSP24 and human GPR45."
RL Biochem. Biophys. Res. Commun. 276:952-956(2000).
CC -!- FUNCTION: Orphan receptor. May play a role in brain function.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: Brain-specific.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF139643; AAC42573.1; -.
DR MGD; MGI:2135884; Gpr63.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm 1; 1.
DR PRINTS; PR00237; GPCR_RHODOPSIN.
DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; FALSE_NEG.
DR PROSITE; PS00262; G_PROTEIN_RECEP_F1_2; 1. Glycoprotein.
KW G-protein coupled receptor; Transmembrane; Glycoprotein.
FT DOMAIN 1 87 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 88 112 1 (POTENTIAL).
FT DOMAIN 113 123 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 124 148 2 (POTENTIAL).
FT DOMAIN 149 165 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 166 190 3 (POTENTIAL).
FT DOMAIN 191 202 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 203 222 4 (POTENTIAL).
FT DOMAIN 223 247 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 248 272 5 (POTENTIAL).
FT DOMAIN 273 321 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 322 345 6 (POTENTIAL).
FT DOMAIN 346 357 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 358 379 7 (POTENTIAL).
FT DOMAIN 380 425 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 34 34 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 68 68 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 425 AA; 47718 MW; B9E99A5521DDED36 CRC64;
```

Query Match 74.4%; Score 32; DB 1; Length 425;
Best Local Similarity 66.7%; Pred. No. 55;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:42:36 ; Search time 25.9695 Seconds
(Without alignments)
89.431 Million cell updates/sec

Title: US-10-089-549-1
Perfect score: 43
Sequence: 1 YVSSLGNI 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL 23.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phage.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_rvivirus.*
 - 16: sp_bacteriophage.*
 - 17: sp_archaea.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	100.0	1284	2 Q8RQU9	Q8RQU9 bacillus gl
2	35	81.4	148	17 O59589	O59589 pyrococcus
3	35	81.4	988	5 O17434	O17434 nephila cla
4	35	81.4	2119	2 Q9AHT5	Q9AHT5 streptococ
5	35	81.4	2140	16 Q97RY6	Q97RY6 streptococ
6	35	81.4	2144	2 Q9S4M8	Q9S4M8 streptococ
7	35	81.4	2144	16 Q8DQP7	Q8DQP7 streptococ
8	34	79.1	539	5 Q8IBS3	Q8IBS3 plasmodium
9	33	76.7	153	16 Q9WU04	Q9WU04 thermotoga
10	33	76.7	304	16 Q8PLC8	Q8PLC8 xanthomonas
11	33	76.7	438	3 P78573	P78573 aspergillus
12	33	76.7	530	8 Q8HMZ6	Q8HMZ6 schizosacch
13	33	76.7	538	2 Q93PC1	Q93PC1 microscilla
14	33	76.7	550	16 Q8D6Q2	Q8D6Q2 vibrio vuln
15	33	76.7	572	2 Q9F7E5	Q9F7E5 acinetobact
16	33	76.7	574	2 Q93RL4	Q93RL4 acinetobact

17	33	76.7	966	16 Q92KQ3	Q92KQ3 rhizobium m
18	33	76.7	1018	16 Q8RFH6	Q8RFH6 fusobacteri
19	33	76.7	4524	5 Q813J9	Q813J9 plasmodium
20	32	74.4	52	11 Q62368	Q62368 mus musculu
21	32	74.4	218	13 Q8JHHS	Q8JHHS brachydanio
22	32	74.4	271	4 Q8N842	Q8N842 homo sapien
23	32	74.4	282	10 Q8GW91	Q8GW91 arabidopsis
24	32	74.4	308	16 Q8XZH2	Q8XZH2 ralstonia s
25	32	74.4	332	4 Q8G9S2	Q8G9S2 homo sapien
26	32	74.4	332	4 Q8NG52	Q8NG52 homo sapien
27	32	74.4	334	16 Q9WZ99	Q9WZ99 thermotoga
28	32	74.4	339	16 Q8RCI6	Q8RCI6 thermoanaer
29	32	74.4	356	16 Q9JZA3	Q9JZA3 neisseria m
30	32	74.4	356	16 Q9JUC6	Q9JUC6 neisseria m
31	32	74.4	387	4 Q8WU13	Q8WU13 homo sapien
32	32	74.4	425	11 Q8BZ93	Q8BZ93 mus musculu
33	32	74.4	486	6 Q9TUB6	Q9TUB6 sus scrofa
34	32	74.4	551	16 Q9HZP5	Q9HZP5 pseudomonas
35	32	74.4	561	16 Q8ZQL6	Q8ZQL6 salmonella
36	32	74.4	568	17 Q8ZUJ4	Q8ZUJ4 pyrobaculum
37	32	74.4	607	2 Q45721	Q45721 bacillus th
38	32	74.4	614	2 Q9L4V3	Q9L4V3 psychrobact
39	32	74.4	618	2 Q45737	Q45737 bacillus th
40	32	74.4	618	2 Q32306	Q32306 bacillus th
41	32	74.4	640	12 Q8B5W8	Q8B5W8 human papil
42	32	74.4	670	17 Q8TKD1	Q8TKD1 methanosarc
43	32	74.4	723	2 Q84AB5	Q84AB5 bacillus th
44	32	74.4	730	16 Q9RS52	Q9RS52 deinococcus
45	32	74.4	761	2 Q8VUM1	Q8VUM1 alteromonas

ALIGNMENTS

RESULT 1

Q8RQU9 PRELIMINARY; PRT; 1284 AA.
AC Q8RQU9;
DT 01-JUN-2002 (TREMREL. 21, Created)
DT 01-JUN-2002 (TREMREL. 21, Last sequence update)
DT 01-MAR-2003 (TREMREL. 23, Last annotation update)
DE 6-glucosyltransferase.
GN CTSZ.
OS Bacillus globisporus.
OC Bacteria; Firmicutes; Bacillales; Planococcaceae; Sporosarcina.
OX NCBI_taxid=1459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C11;
RA Maruta K.;
RT *Cloning and sequencing of the genes encoding cyclic tetrasaccharide-
synthesizing enzymes from Bacillus globisporus C11";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB073929; BAB88404.1; -;
DR InterPro; IPR005084; CEM 6.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF03422; CEM_6; 2.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Transferase.
SQ SEQUENCE 1284 AA; 139128 MW; 2167B0DE84F42E9C CRC64;

Query Match 100.0%; Score 43; DB 2; Length 1284;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YVSSLGNI 9

Db 36 YVSSLGNI 44

RESULT 2

Q9AHT5
ID Q9AHT5 PRELIMINARY; PRT; 2119 AA.
AC Q9AHT5;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Serine protease (Fragment).
GN PRGA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N4;
RX MEDLINE=21116976; PubMed=11179332;
RA Wiemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,
Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,
Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,
Langermann S., Johnson S., Koenig S.,
RT "Use of a Whole Genome Approach to Identify Vaccine Molecules
Affording Protection against Streptococcus pneumoniae Infection.";
RL Infect. Immun. 69:1593-1598(2001).
CC - SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY
AN AMIDE BOND (BY SIMILARITY).
DR EMBL; AF291699; AAK19159.1; -;
DR HSSP; P00782; 2SPT.
DR MEROPS; S08.064; -;
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR006192; LPXTG.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 2.
DR PRINTS; PR00723; SUBTILISIN.
DR TIGRFA; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00840; PA; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR PROSITE; PS00678; WD_REPEATS; 1; 1.
XM Cell wall; Peptidoglycan-anchor; Protease.
FT NON TER 1
SQ SEQUENCE 2119 AA; 238226 MW; 517F9B7F6B960A6A CRC64;
Query Match 81.4%; Score 35; DB 2; Length 2119;
Best Local Similarity 77.8%; Pred. No. 4e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 YVSSLGNI 9
Db 1534 YVSSYGNVI 1542
RESULT 5
Q97RY6 PRELIMINARY; PRT; 2140 AA.
ID Q97RY6
AC Q97RY6;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Serine protease, subtilase family.
GN SP0641.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463916;

O59589 PRELIMINARY; PRT; 148 AA.
AC O59589;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein PH1926.
GN PH1926.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kishida K., Oguchi A.,
Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
Mauchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR EMBL; AP000007; BAA31053.1; -;
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 148 AA; 17911 MW; 810236EAC3BPSF2F CRC64;
Query Match 81.4%; Score 35; DB 17; Length 148;
Best Local Similarity 55.6%; Pred. No. 22;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 YVSSLGNI 9
Db 100 YIDSIGNLV 108
RESULT 3
O17434 PRELIMINARY; PRT; 988 AA.
ID O17434
AC O17434;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Minor ampullate silk protein M1Sp1 (Fragment).
OS Nephila clavipes (Orb spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
OX NCBI_TaxID=6915;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98200471; PubMed=9541398;
RA Colgin M.A., Lewis R.V.;
RT "Spider minor ampullate silk proteins contain new repetitive sequences
and highly conserved non-silk-like 'spacer regions'.";
RL Protein Sci. 7:667-672(1998).
DR EMBL; AF027735; AAC14589.1; -;
DR InterPro; IPR000817; PRION.
DR PRINTS; PR00341; PRION.
FT NON TER 1
SQ SEQUENCE 988 AA; 79082 MW; 461E03DF53F7085D CRC64;
Query Match 81.4%; Score 35; DB 5; Length 988;
Best Local Similarity 77.8%; Pred. No. 1.7e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 YVSSLGNI 9
Db 152 YVSLGNAL 160
RESULT 4

Wed Jan 21 11:28:18 2004

us-10-089-549-1.rsp

RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
 RA Holtzapple E., Khouli H., Wolf A.M., Utterback T.R., Hansen C.L., E.K.,
 RA McDonald L.A., Feldblyum T.V., Anguoli S., Dickinson T., Hickey E.K.,
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RT "Complete genome sequence of a virulent isolate of *Streptococcus*
 RT *pneumoniae*."; 498-506 (2001).
 RL Science 293:498-506 (2001).
 DR EMBL; AF007373; AAK74791.1; -;
 DR MEROPS; S08.064; -;
 DR TIGR; SP0641; -;
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR InterPro; IPR006192; LPXTG.
 DR InterPro; IPR003137; PA.
 DR InterPro; IPR003209; Peptidase_S8.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00746; Gram_pos_anchor; 1.
 DR Pfam; PF02225; PA; 1.
 DR Pfam; PF00082; Peptidase_S8; 2.
 DR PRINTS; PRO0723; SUBTILISIN.
 DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
 DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
 DR PROSITE; PS00840; PA; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 DR PROSITE; PS00678; WD_REPEATS_1; 1.
 KW Pfam; PF00746; Gram_pos_anchor; 1.
 KW Protease; Complete proteome.
 SQ SEQUENCE 2140 AA; 240426 MW; FA44AD8E29385334 CRC64;
 Query Match 81.4%; Score 35; DB 16; Length 2140;
 Best Local Similarity 77.8%; Pred. No. 4e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 YVSSLGNI 9
 DB 1555 YVSSYGNVI 1563
 RESULT 6
 Q9S4M8 PRELIMINARY; PRT; 2144 AA.
 AC Q9S4M8
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Cell wall-associated serine proteinase PrtA precursor.
 GN PrtA.
 OS *Streptococcus pneumoniae*.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 CX NCBI_TaxID=1313;
 RN SEQUENCE FROM N.A.
 RP STRAIN=3.B;
 RC STRAIN=3.B;
 RA Bethé G., ten Thoren E., Bongaerts R.J.M., Heinz H.-P., Zysk G.;
 RT "Cloning and sequencing of a novel surface protease of *Streptococcus*
 RT *pneumoniae*."; 1-11 (2000).
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY
 CC AN AMIDE BOND (BY SIMILARITY).
 DR EMBL; AF127143; AAD48399.1; -;
 DR HSSP; P00782; 2SST.
 DR MEROPS; S08.064; -;
 DR InterPro; IPR001893; Gram_pos_anchor.
 DR InterPro; IPR006192; LPXTG.
 DR InterPro; IPR003137; PA.
 DR InterPro; IPR003209; Peptidase_S8.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00746; Gram_pos_anchor; 1.
 DR Pfam; PF02225; PA; 1.

DR Pfam; PF00082; Peptidase_S8; 2.
 DR PRINTS; PRO0723; SUBTILISIN.
 DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
 DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
 DR PROSITE; PS00840; PA; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 DR PROSITE; PS00678; WD_REPEATS_1; 1.
 KW Cell wall; Peptidoglycan-anchor; Signal.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 2144 CELL WALL-ASSOCIATED SERINE PROTEINASE
 FT PrtA.
 SQ SEQUENCE 2144 AA; 240724 MW; 2052511470741331 CRC64;
 Query Match 81.4%; Score 35; DB 2; Length 2144;
 Best Local Similarity 77.8%; Pred. No. 4e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 YVSSLGNI 9
 DB 1559 YVSSYGNVI 1567
 RESULT 7
 Q8DQ7 PRELIMINARY; PRT; 2144 AA.
 AC Q8DQ7
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Cell wall-associated serine proteinase PrtA (EC 3.4.21.-).
 GN PrtA OR SPR0561.
 OS *Streptococcus pneumoniae* (strain ATCC BAA-255 / R6).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 CX NCBI_TaxID=171101;
 RN SEQUENCE FROM N.A.
 RP MEDLINE=21429245; PubMed=11544234;
 RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
 RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
 RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
 RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
 RA Mahren S.M., McHenry M., McLeaster K., Mundy C.W., Nicas T.I.,
 RA Norris F.H., O'Garra M., Peery R.B., Robertson G.T., Rocky P.,
 RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
 RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
 RA Glass J.I.;
 RT "Genome of the bacterium *Streptococcus pneumoniae* strain R6.";
 RL J. Bacteriol. 183:5709-5717 (2001).
 DR EMBL; AB008434; AAK93365.1; -;
 DR Hydrolyase; Complete proteome.
 SQ SEQUENCE 2144 AA; 240436 MW; 8C1B4B1DBC503A0C CRC64;
 Query Match 81.4%; Score 35; DB 16; Length 2144;
 Best Local Similarity 77.8%; Pred. No. 4e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 YVSSLGNI 9
 DB 1559 YVSSYGNVI 1567
 RESULT 8
 Q8IBS3 PRELIMINARY; PRT; 539 AA.
 AC Q8IBS3
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Seryl-trna synthetase, putative (EC 6.1.1.11).
 GN PF07_0073.
 OS *Plasmodium falciparum* (isolate 3D7).

```

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
RA Quail M., Barrell B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RW EXBL: AL844506; CAD50925.1; -.
KW Aminoacyl-tRNA synthetase; Ligase.
SQ SEQUENCE 539 AA; 62454 MW; F29DAC5F3C14388A CRC64;

Query Match 79.1%; Score 34; DB 5; Length 53;
Best Local Similarity 55.6%; Pred. No. 1.5e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 YVSSLGNLI 9
Db 107 YISKIGNLL 115

RESULT 9
Q9WXU4 PRELIMINARY; PRT; 153 AA.
ID Q9WXU4
AC Q9WXU4
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein TM0089.
GN TM0089.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=23336;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton V.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RA "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima."
RL Nature 399:323-329(1999).
DR EMBL; AE001695; AAD35183.1; -.
DR TIGR; TM0089; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 153 AA; 17143 MW; 3059E8A5E96C12A1 CRC64;

Query Match 76.7%; Score 33; DB 16; Length 153;
Best Local Similarity 77.8%; Pred. No. 61;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YVSSLGNLI 9
Db 133 YVSSLGLV 141

RESULT 10
Q8PLC8 PRELIMINARY; PRT; 304 AA.
ID Q8PLC8
AC Q8PLC8
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Regulatory protein.
GN RPF OR XAC1874.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Parah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida M.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Canarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RL host specificities."
RL Nature 417:459-463(2002).
DR EMBL; AE011820; AAM36736.1; -.
KW Complete proteome.
SQ SEQUENCE 304 AA; 33767 MW; 89278C973F657D44 CRC64;

Query Match 76.7%; Score 33; DB 16; Length 304;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 YVSSLGNLI 8
Db 32 YVSLGNV 39

RESULT 11
P78573 PRELIMINARY; PRT; 438 AA.
ID P78573
AC P78573
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Fructosyl amine: oxygen oxidoreductase.
OS Aspergillus fumigatus (Sartorya fumigata).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5085;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97284723; PubMed=9139700;
RA Takahashi M., Pischetrieder M., Monnier V.M.;
RT "Molecular cloning and expression of amadoriase isoenzyme (fructosyl
RL amine:oxygen oxidoreductase, EC 1.5.3) from Aspergillus fumigatus."
RL J. Biol. Chem. 272:12505-12507(1997).
DR EMBL; U82830; AAC49711.1; -.
DR InterPro; IPR00205; NAD binding.
DR InterPro; IPR000594; Thif domain.
SQ SEQUENCE 438 AA; 48931 MW; FBC8A3E5D89D02AE CRC64;

Query Match 76.7%; Score 33; DB 3; Length 438;
Best Local Similarity 66.7%; Pred. No. 1.9e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 YVSSLGNLI 9
Db 369 YLFSIGNLI 377

RESULT 12
Q8HMZ6 PRELIMINARY; PRT; 530 AA.
ID Q8HMZ6
AC Q8HMZ6
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

```

Wed Jan 21 11:28:18 2004

us-10-089-549-1.rapt

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Cytochrome c oxidase subunit 1 (EC 1.9.3.1).
 GN COX1.
 OS Schizosaccharomyces japonicus.
 OG Mitochondrion.
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4897;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Billerwell C.B.; Leigh J.; Forget L.; Lang F.B.F.;
 RT "A Comparison of Three Fission Yeast Mitochondrial Genomes."
 RL Nucleic Acids Res. 0:0-0(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Lang F.B.F.;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF547983; AAN37913.1; -.
 KW Oxidoreductase; Mitochondrion.
 SQ SEQUENCE 530 AA; 58822 MW; 5DD2960A6C0A1CB2 CRC64;

Query Match 76.7%; Score 33; DB 8; Length 530;
 Best Local Similarity 55.6%; Pred. No. 2.3e+02; Indels 0; Gaps 0;
 Matches 5; Conservative 4; Mismatches 0;

QY 1 YVSSLGNLI 9
 DB 452 YVSSIGSMI 460
 |||||:|:|

RESULT 13
 Q93PC1 PRELIMINARY; PRT; 538 AA.
 AC Q93PC1; (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE MS108.
 OS Microscilla sp. PRE1.
 OG Plasmid pSBL5.
 OC Bacteria; Bacteroidetes; Sphingobacteria; Sphingobacteriales;
 OC Flexibacteraceae; Microscilla.
 OX NCBI_TaxID=155537;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PRE1;
 RA Zhong Z.; Toukdarian A.; Helinski D.; Knauf V.; Sykes S.;
 RA Wilkinson J.B.; O'Bryne C.; Shea T.; De Loughery C.; Caspi R.;
 RT "Sequence Analysis of a 101 kb Plasmid Required for Agar-degradation
 of a Microscilla strain."
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF339846; AAK62830.1; -.
 KW Plasmid.
 SQ SEQUENCE 538 AA; 58501 MW; 460E2DD75B0FA321 CRC64;

Query Match 76.7%; Score 33; DB 2; Length 538;
 Best Local Similarity 75.0%; Pred. No. 2.4e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YVSSLGNLI 8
 DB 446 YVESLGNL 453
 |||||:|:|

RESULT 14
 Q8D6Q2 PRELIMINARY; PRT; 550 AA.
 AC Q8D6Q2;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Dehydrogenase.

GN VV20476.
 OS Vibrio vulnificus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=672;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CMCP6;
 RA Rhee J.H.; Kim S.Y.; Chung S.S.; Kim J.J.; Moon Y.H.; Jeong H.;
 RA Choy H.E.;
 RT "Complete genome sequence of Vibrio vulnificus CMCP6."
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB016809; AAO07427.1; -.
 KW Complete proteome.
 SQ SEQUENCE 550 AA; 59924 MW; 82BC7C3EC4C04917 CRC64;

Query Match 76.7%; Score 33; DB 16; Length 550;
 Best Local Similarity 87.5%; Pred. No. 2.4e+02;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YVSSLGNLI 8
 DB 109 YVSSLGNL 116
 |||||:|:|

RESULT 15
 Q9F7E5 PRELIMINARY; PRT; 572 AA.
 AC Q9F7E5;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE Ubiquinone oxidoreductase.
 GN ETEF.
 OS Acinetobacter sp. SE19.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Moraxellaceae; Acinetobacter.
 OX NCBI_TaxID=135835;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SE19;
 RX MEDLINE=20398154; PubMed=10940013;
 RA Cheng Q.; Thomas S.M.; Kostichka K.; Valentine J.R.; Nagarajan V.;
 RT "Genetic analysis of a gene cluster for cyclohexanol oxidation in
 acinetobacter sp. strain SE19 by in vitro transposition."
 RL J. Bacteriol. 182:4744-4751(2000).
 DR EMBL; AF282240; AAG10020.1; -.
 DR InterPro; IPR003042; Rng_mnxygenase.
 DR PRINTS; PR00420; RNMNMKNASE.
 SQ SEQUENCE 572 AA; 63149 MW; 6C47BCE06F8742F0 CRC64;

Query Match 76.7%; Score 33; DB 2; Length 572;
 Best Local Similarity 66.7%; Pred. No. 2.5e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 YVSSLGNLI 9
 DB 119 YVISLGNV 127
 |||||:|:|

Search completed: January 20, 2004, 17:00:31
 Job time : 27.9695 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:41:26 ; Search time 48.8244 Seconds
(without alignments)
42.263 Million cell updates/sec

Title: US-10-089-549-18

Perfect score: 65

Sequence: 1 APLGVQRAQFQSG 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 19Jun03:*

```
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	65	100.0	13	ABG30557	Alpha-isomaltosylg
2	65	100.0	13	AAO17347	Alpha-isomaltosylg
3	65	100.0	929	ABG30539	Alpha-isomaltosylg
4	65	100.0	965	ABG30565	Alpha-isomaltosylg
5	39	60.0	118	ABP40427	Staphylococcus epi
6	39	60.0	227	AAAB47301	Glycoprotein E-RNS
7	39	60.0	228	AAW931198	Pestivirus Erns cy
8	39	60.0	228	AAV82117	Bovine viral diarr
9	39	60.0	228	AAV82121	Classical swine fe

10	39	60.0	271	20	AAW931195	Pestivirus Erns cy
11	39	60.0	271	20	AAW931196	Pestivirus Erns cy
12	39	60.0	271	20	AAW931197	Pestivirus Erns cy
13	39	60.0	453	13	AAW29511	BDV structural gly
14	39	60.0	731	13	AAW29512	BDV structural gly
15	39	60.0	823	13	AAW29510	BDV structural gly
16	39	60.0	1069	13	AAW29513	Fusion of BDV stru
17	39	60.0	1286	8	AAW70340	Bovine diarrhoea v
18	39	60.0	1536	22	ABG22570	Novel human diagno
19	39	60.0	3896	21	AAV53617	Amino acid sequenc
20	39	60.0	3988	21	AAV53615	NADL protein encod
21	39	60.0	3988	21	AAV53616	Amino acid sequenc
22	37	56.9	119	22	AAW382139	S. epidermidis ope
23	37	56.9	897	23	AAU74766	Bovine viral diarr
24	36.5	56.2	1134	19	AAW46859	Bacillus thuringie
25	36.5	56.2	1134	20	AAW24962	Bacillus thuringie
26	36	55.4	417	21	AAW41242	Arabidopsis thalia
27	36	55.4	422	21	AAW41241	Arabidopsis thalia
28	36	55.4	699	22	ABW71219	Drosophila melanog
29	36	55.4	1199	22	ABW71217	Drosophila melanog
30	36	55.4	1408	22	ABG22573	Novel human diagno
31	35	53.8	18	15	AAW52588	Hepatitis E virus
32	35	53.8	18	16	AAW82209	HEV ORF2 epitope (
33	35	53.8	42	16	AAW71106	HEV antigenic pept
34	35	53.8	42	18	AAW35822	Hepatitis E virus
35	35	53.8	42	19	AAW80204	ET-NANB (HEV) Burm
36	35	53.8	42	19	AAW71217	Epitope of the Bur
37	35	53.8	42	21	AAW24127	Hepatitis E virus
38	35	53.8	42	22	AAW62530	HEV-Burma strain d
39	35	53.8	42	23	AAO15706	Hepatitis E virus
40	35	53.8	48	17	AAW96095	HEV US-1 peptide e
41	35	53.8	48	20	AAW311390	Drosophila melanog
42	35	53.8	77	22	ABW68810	Novel human diagno
43	35	53.8	89	22	ABG00731	Hepatitis E virus
44	35	53.8	176	16	AAW82201	Hepatitis E virus
45	35	53.8	267	23	ABW04831	Human hepatitis E

ALIGNMENTS

RESULT 1
ABG30557
ID ABG30557 standard; Peptide; 13 AA.

AC ABG30557;

XX 07-OCT-2002 (first entry)

XX Alpha-isomaltosylglucosaccharide synthase related peptide #18.

XX Alpha-isomaltosylglucosaccharide synthase; sugar production;
XX cyclic tetrasaccharide; cyclo(-6)-alpha-D-glucopyranosyl(1-3);
XX alpha-isomaltosyltransferase; viscosity modifier; sweetener;
XX heat stabiliser; hydrolysis stabiliser; colour stabiliser; foodstuff;
XX cosmetic; drug composition.

XX Unidentified.

XX WO200255708-A1.

XX 18-JUL-2002.

XX 09-JAN-2002; 2002WO-JP000052.

XX 12-JAN-2001; 2001JP-0005441.

XX (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.

XX Kubota M, Maruta K, Yamamoto T, Fukuda S;

XX MPI; 2002-520129/55.

PT New alpha-isomaltosylglucosaccharide synthase of bacterial origin for
 PT the production of cyclic tetrasaccharide gum -
 XX Disclosure; Page 135; 144pp; Japanese.
 XX
 CC The invention describes novel microbial polypeptides having
 CC alpha-isomaltosylglucosaccharide synthase activity. The proteins are
 CC useful for producing a sugar (I) having at least three glucose units
 CC connected by alpha-1,4 bonds except for alpha-1,6 at the non-reducing
 CC end. The invention also describes a method for the production of the
 CC cyclic tetrasaccharide cyclo(-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-
 CC glucopyranosyl(1-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-glucopyranosyl
 CC (I-) (II) by treatment of (I) with alpha-isomaltosyltransferase.
 CC (II) and similar sugars in crystalline or syrup form are used as
 CC a viscosity modifier, sweetener, heat stabiliser, hydrolysis stabiliser
 CC and colour stabiliser for foodstuffs, cosmetics and drug compositions.
 CC This is the amino acid sequence of an alpha-isomaltosylglucosaccharide
 CC synthase related peptide.
 XX
 SQ Sequence 13 AA;
 Query Match 100.0%; Score 65; DB 23; Length 13;
 Best Local Similarity 100.0%; Pred. No. 3.7e-06;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 APLGVQRAQFQSG 13
 |||||
 DB 1 APLGVQRAQFQSG 13
 |||||
 RESULT 2
 AA017347
 ID AA017347 standard; peptide; 13 AA.
 XX
 AC AA017347;
 XX
 DT 08-JUL-2002 (first entry)
 XX
 DE Alpha-isomaltosylglucosaccharide synthase related peptide #18.
 XX
 KW Alpha-isomaltosylglucosaccharide synthase; enzyme; sweetener; food;
 KW drink; cosmetics; pharmaceutical; cyclic tetrasaccharide.
 XX
 OS Arthrobacter globiformis.
 XX
 PN WO200210361-A1.
 XX
 PD 07-FEB-2002.
 XX
 PF 25-JUL-2001; 2001WO-JP06412.
 XX
 PR 01-AUG-2000; 2000JP-0233364.
 XX
 PR 02-AUG-2000; 2000JP-0234937.
 XX
 PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.
 XX
 PI Kubota M., Tausaki K., Higashiyama T., Fukuda S., Miyake T;
 XX
 XX WPI; 2002-315302/35.
 XX
 PT Alpha-isomaltosylglucosaccharide synthase capable of transferring
 PT Alpha-glucosyl from saccharide, useful in producing cyclic
 PT tetrasaccharides and related carbohydrates industrially for application
 PT in compositions e.g. as drugs -
 XX
 XX Claim 5; Page 204; 209pp; Japanese.
 XX
 CC The present invention relates to an alpha-isomaltosylglucosaccharide
 CC synthase which is capable of transferring alpha-glucosyl from a
 CC saccharide to form a specific saccharide which carries an alpha-1,6
 CC glucosyl linkage at its non-reducing end and an alpha-1,4 glucosyl
 CC linkage at ends other than the non-reducing end and has a degree of
 CC glucose polymerization of at most 3, but without substantially elevating

CC the reducing ability. The synthase is useful in producing cyclic
 CC tetrasaccharides and related carbohydrates industrially for application
 CC in compositions e.g. as sweeteners, low calorie food materials, taste
 CC improvers, flavour improvers, quality improvers, water-separation
 CC inhibitors, stabilizers, excipients, binding agents and pulverisation
 CC bases, especially in foods, drinks, cosmetics and pharmaceuticals. The
 CC present sequence is a peptide described in the exemplification of the
 CC invention.
 XX
 SQ Sequence 13 AA;
 Query Match 100.0%; Score 65; DB 23; Length 13;
 Best Local Similarity 100.0%; Pred. No. 3.7e-06;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 APLGVQRAQFQSG 13
 |||||
 DB 1 APLGVQRAQFQSG 13
 |||||
 RESULT 3
 ABG30539
 ID ABG30539 standard; Protein; 929 AA.
 XX
 AC ABG30539;
 XX
 DT 07-OCT-2002 (first entry)
 XX
 DE Alpha-isomaltosylglucosaccharide synthase #3 mature protein.
 XX
 KW Alpha-isomaltosylglucosaccharide synthase; sugar production;
 KW cyclic tetrasaccharide; cyclo(-6)-alpha-D-glucopyranosyl(1-3);
 KW alpha-isomaltosyltransferase; viscosity modifier; sweetener;
 KW heat stabiliser; hydrolysis stabiliser; colour stabiliser; foodstuff;
 KW cosmetic; drug composition; enzyme.
 XX
 OS Unidentified.
 XX
 PN WO200255708-A1.
 XX
 PD 18-JUL-2002.
 XX
 PF 09-JAN-2002; 2002WO-JP00052.
 XX
 PR 12-JAN-2001; 2001JP-0005441.
 XX
 PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.
 XX
 PI Kubota M., Maruta K., Yamamoto T., Fukuda S;
 XX
 XX WPI; 2002-520129/55.
 XX
 DR N-PSDB; ABK88156.
 XX
 PT New alpha-isomaltosylglucosaccharide synthase of bacterial origin for
 PT the production of cyclic tetrasaccharide gum -
 XX
 XX Claim 1; Page 109-112; 144pp; Japanese.
 XX
 CC The invention describes novel microbial polypeptides having
 CC alpha-isomaltosylglucosaccharide synthase activity. The proteins are
 CC useful for producing a sugar (I) having at least three glucose units
 CC connected by alpha-1,4 bonds except for alpha-1,6 at the non-reducing
 CC end. The invention also describes a method for the production of the
 CC cyclic tetrasaccharide cyclo(-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-
 CC glucopyranosyl(1-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-glucopyranosyl
 CC (I-) (II) by treatment of (I) with alpha-isomaltosyltransferase.
 CC (II) and similar sugars in crystalline or syrup form are used as
 CC a viscosity modifier, sweetener, heat stabiliser, hydrolysis stabiliser
 CC and colour stabiliser for foodstuffs, cosmetics and drug compositions.
 CC This is the amino acid sequence of an alpha-isomaltosylglucosaccharide
 CC synthase mature protein.
 XX
 XX Sequence 929 AA;

Query Match 100.0%; Score 65; DB 23; Length 923;
Best Local Similarity 100.0%; Pred. No. 0.00036;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 APLGVQRAQFQSG 13
Db 1 APLGVQRAQFQSG 13

RESULT 4
ABG30565
ID ABG30565 standard; Protein; 965 AA.
XX AC ABG30565;
XX DT 07-OCT-2002 (first entry)
XX DE Alpha-isomaltosylglucosaccharide synthase #3.
XX KW Alpha-isomaltosylglucosaccharide synthase; sugar production;
XX KW cyclic tetrasaccharide; cyclo(-6)-alpha-D-glucopyranosyl(1-3);
XX KW alpha-isomaltosyltransferase; viscosity modifier; sweetener;
XX KW heat stabiliser; hydrolysis stabiliser; colour stabiliser; foodstuff;
XX KW cosmetic; drug composition.
XX OS Unidentified.
XX PN WO200255708-A1.
XX PD 18-JUL-2002.
XX PF 09-JAN-2002; 2002WO-JP00052.
XX PR 12-JAN-2001; 2001JP-0005441.
XX PA (HAYB) HAYASHIBARA SEIITSU KAGAKU.
XX PI Kubota M, Maruta K, Yamamoto T, Fukuda S;
XX WPI; 2002-520129/55.
XX PT New alpha-isomaltosylglucosaccharide synthase of bacterial origin for
XX PT the production of cyclic tetrasaccharide gum -
XX PS Disclosure; Page 137-141; 144pp; Japanese.
XX CC The invention describes novel microbial polypeptides having
XX CC alpha-isomaltosylglucosaccharide synthase activity. The proteins are
XX CC useful for producing a sugar (I) having at least three glucose units
XX CC connected by alpha-1,4 bonds except for alpha-1,6 at the non-reducing
XX CC end. The invention also describes a method for the production of the
XX CC cyclic tetrasaccharide cyclo(-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-
XX CC glucopyranosyl(1-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-glucopyranosyl
XX CC (1-1) (II) by treatment of (I) with alpha-isomaltosyltransferase.
XX CC (II) and similar sugars in crystalline or syrup form are used as
XX CC a viscosity modifier, sweetener, heat stabiliser, hydrolysis stabiliser
XX CC and colour stabiliser for foodstuffs, cosmetics and drug compositions.
XX CC This is the amino acid sequence of an alpha-isomaltosylglucosaccharide
XX CC synthase.
XX SQ Sequence 965 AA;

Query Match 100.0%; Score 65; DB 23; Length 965;
Best Local Similarity 100.0%; Pred. No. 0.00037;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 APLGVQRAQFQSG 13
Db 37 APLGVQRAQFQSG 49

RESULT 5

ABP40427
ID ABP40427 standard; Protein; 118 AA.
XX AC ABP40427;
XX DT 24-JUL-2002 (first entry)
XX DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5272.
XX KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX KW antibacterial; gene therapy.
XX OS Staphylococcus epidermidis.
XX PN US6380370-B1.
XX PD 30-APR-2002.
XX PF 13-AUG-1998; 98US-0134001.
XX PR 14-AUG-1997; 97US-055779P.
XX PR 08-NOV-1997; 97US-064964P.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Doucette-Stamm LA, Bush D;
XX WPI; 2002-381255/41.
XX N-PSDB; ABN92972.
XX PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
XX PT polypeptide, useful for diagnosing and treating bacterial infections -
XX PS Disclosure; SEQ ID 5272; 267pp; English.
XX CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
XX CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
XX CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
XX CC antibacterial activity and can be used in gene therapy. The sequences
XX CC can also be used in the diagnosis and treatment of bacterial infections,
XX CC particularly S. epidermidis infections. The sequences can be used to
XX CC screen for compounds able to interfere with the S. epidermidis life
XX CC cycle or inhibit S. epidermidis infection.
XX CC N.B. The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from the
XX CC USPTO web site.
XX SQ Sequence 118 AA;

Query Match 60.0%; Score 39; DB 23; Length 118;
Best Local Similarity 58.3%; Pred. No. 6.7;
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 APLGVQRAQFQSG 12
Db 49 APLGVRRSEWDS 60

RESULT 6
AAB47301
ID AAB47301 standard; protein; 227 AA.
XX AC AAB47301;
XX DT 29-AUG-2001 (first entry)
XX DE Glycoprotein E-RNS of BVDV CP VII.
XX KW Glycoprotein E-RNS; bovine viral diarrhoea virus; BVDV; coat protein;
XX KW CP 7; RNase activity; vaccine; breeding stock; cattle; fetus.
XX OS Bovine viral diarrhoea virus.

FH Key Location/Qualifiers
 FT Region 28..40
 FT /note= "Conserved region w.r.t RNase"
 FT Region 71..90
 FT /note= "Conserved region w.r.t RNase"
 FT Active-site 30
 FT /note= "Catalytic centre"
 FT Active-site 79
 FT /note= "Catalytic centre, deleted to inactivate virus"
 XX
 FN WO200139801-A2.
 XX 07-JUN-2001.
 XX
 XX 29-NOV-2000; 2000WO-EP11940.
 XX
 XX 30-NOV-1999; 99EP-0123767.
 XX 14-DEC-1999; 99US-C170616.
 XX (BOEH) BOEHRINGER INGELHEIM VETMEDICA GMBH.
 XX Meyers G, Elbers K;
 XX WPI; 2001-408259/43.
 XX
 XX Use of live bovine diarrhoea virus with inactivated RNase activity, for
 PT the preparation of live vaccine for use in the prevention and/or
 PT treatment of BVDV infections in breeding stocks of cattle -
 XX
 XX Claim 7; Fig 1; 21pp; English.
 XX
 XX This sequence represents the amino acid sequence of the glycoprotein
 CC E-RNS of bovine viral diarrhoea virus (BVDV). This sequence corresponds
 CC to residues 271-497 of the CP 7 protein. The RNase activity residing in
 CC this glycoprotein is inactivated by deletion of the Histidine
 CC residue at position 79. The inactivated strain may be used for
 CC the preparation of a live vaccine for use in the prevention and/or
 CC treatment of BVDV infections in breeding stocks of cattle. The vaccine
 CC composition is suitable for preventing and protecting the fetus
 CC itself against BVDV with live vaccine, and is safe.
 XX
 SQ Sequence 227 AA;
 Query Match 60.0%; Score 39; DB 22; Length 227;
 Best Local Similarity 70.0%; Pred. No. 13;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 4 GVQRAQFQSG 13
 Db 15 GIQRAQFQSG 24
 RESULT 7
 AAW93198
 ID AAW93198 standard; protein; 228 AA.
 XX
 AC AAW93198;
 XX
 XX 08-JUN-1999 (first entry)
 XX
 XX Pestivirus Ems cytotoxic polypeptide fragment pEms-bvd27.
 XX
 XX Cytotoxic; Ems; ribonuclease activity; pharmaceutical composition;
 KW therapy; medicament; lymphocyte; rheumatism; multiple sclerosis;
 KW organ transplantation; treatment; side effect; anti-fertility;
 KW ribosome-inactivating.
 XX
 OS Pestivirus.
 XX
 XX WO9906569-A1.
 XX
 XX 11-FEB-1999.
 XX

PF 29-JUL-1998; 98WO-NL00434.
 XX
 PR 30-JUL-1997; 97EP-0202371.
 XX
 PA (DIER-) STICHTING INST DIERHOUDERIJ EN DIERGEZON.
 XX
 PI Bruschke CJM, Van Oirschot JT, Van Rijn PA;
 XX WPI; 1999-153798/13.
 XX
 DR New selective cytotoxic polypeptide lacking ribonuclease activity -
 PT useful for therapy and for producing medicaments with a selective
 PT cytotoxic effect on lymphocytes
 XX
 PS Disclosure; Fig 4; 41pp; English.
 XX
 CC This invention describes the identification of novel selective cytotoxic
 CC polypeptides (or fragments) partially or wholly lacking ribonuclease
 CC activity. Such polypeptides are derived from a protein having
 CC ribonuclease activity. The polypeptides form a pharmaceutical
 CC composition, which is useful in therapy, and for the production of a
 CC medicament with a selective cytotoxic effect on lymphocytes. Conditions
 CC treated include rheumatism, multiple sclerosis and post organ
 CC transplantation treatment. The use of the selective agents lacking
 CC ribonuclease activity avoids cytotoxic side effects, including
 CC anti-fertility and ribosome-inactivating effects.
 XX
 SQ Sequence 228 AA;
 Query Match 60.0%; Score 39; DB 20; Length 228;
 Best Local Similarity 70.0%; Pred. No. 14;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 4 GVQRAQFQSG 13
 Db 15 GIQRAQFQSG 24
 RESULT 8
 AAW82117
 ID AAW82117 standard; Protein; 228 AA.
 XX
 AC AAW82117;
 XX
 DT 05-JUN-2000 (first entry)
 XX
 XX Bovine viral diarrhoea virus E-RNS amino acid sequence SEQ ID NO:6.
 DE
 XX Bovine viral diarrhoea virus; BVDV; E2; E-RNS; pestivirus; vaccine;
 KW vaccination; viral disease; infection; immunodominant;
 KW classical swine fever virus; Border disease virus.
 XX
 OS Bovine viral diarrhoea virus.
 XX
 PN EP982402-A1.
 XX
 PD 01-MAR-2000.
 XX
 PF 14-AUG-1998; 98EP-0202737.
 XX
 PR 14-AUG-1998; 98EP-0202737.
 XX
 PA (DIER-) STICHTING INST DIERHOUDERIJ EN DIERGEZON.
 XX
 PI Moormann RJM, Bouma A, Van Rijn PA, De Smit AJ;
 XX WPI; 2000-258615/23.
 DR
 XX Recombinant nucleic acids encoding modified pestiviral proteins for
 PT generating chimeric pestiviral vaccines to protect against classical
 PT swine fever -
 XX
 PS Example 1; Fig 1A; 45pp; English.

XX The present invention describes a recombinant nucleic acid (I) derived
 CC from a pestivirus genotype, where a first fragment (A) encoding at least
 CC one immunodominant part, at least partly responsible for providing
 CC protection against a wildtype pestivirus infection of a viral protein
 CC (P), has been modified to prevent antibodies from being formed against
 CC it. Also described are: (1) a chimeric pestivirus (II) comprising (I);
 CC (2) a vaccine (III) comprising (II); and (3) a method (IV) of controlling
 CC and/or eradicating a wildtype pestivirus infection. (III) may be used
 CC to protect against wildtype pestivirus infection caused by classical
 CC swine fever virus (CSFV), bovine viral diarrhoea virus (BVDV) and Border
 CC disease virus (BDV). When (III) contains Flc7 or Flc9 it may be used as
 CC an emergency vaccination to control outbreaks of CSF when rapid immunity
 CC is required and prophylactic vaccination in areas where classical swine
 CC fever (CSF) is endemic. The present sequence represents a BVDV E-RNS
 CC amino acid sequence given in an example from the present invention.

XX Sequence 228 AA;
 SQ Query Match 60.0%; Score 39; DB 21; Length 228;
 Best Local Similarity 70.0%; Pred. No. 14;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 GVQRAQFQSG 13
 Db 15 GIQRAMFQRG 24
 :||| |||

RESULT 9
 AAY82121
 ID AAY82121 standard; Protein; 228 AA.
 AC AAY82121;
 XX 05-JUN-2000 (first entry)
 DT
 DE Classical swine fever virus E-RNS amino acid sequence SEQ ID NO:10.
 XX
 KW Bovine viral diarrhoea virus; BVDV; E2; E-RNS; pestivirus; vaccine;
 KW vaccination; viral disease; infection; immunodominant;
 KW classical swine fever virus; Border disease virus.
 XX
 OS Hog cholera virus.
 XX
 PN EP982402-A1.
 XX
 PD 01-MAR-2000.
 XX
 PF 14-AUG-1998; 98EP-0202737.
 XX
 PR 14-AUG-1998; 98EP-0202737.
 XX
 PA (DIER-) STICHTING INST DIERHOUDERIJ EN DIERGEZON.
 XX
 PI Moormann RJM, Bouma A, Van Rijn PA, De Smit AJ;
 XX
 DR WPI; 2000-258615/23.
 XX
 PT Recombinant nucleic acids encoding modified pestiviral proteins for
 PT generating chimeric pestiviral vaccines to protect against classical
 PT swine fever -
 XX
 FS Example 1; Fig 1A; 45pp; English.
 XX
 CC The present invention describes a recombinant nucleic acid (I) derived
 CC from a pestivirus genotype, where a first fragment (A) encoding at least
 CC one immunodominant part, at least partly responsible for providing
 CC protection against a wildtype pestivirus infection of a viral protein
 CC (P), has been modified to prevent antibodies from being formed against
 CC it. Also described are: (1) a chimeric pestivirus (II) comprising (I);
 CC (2) a vaccine (III) comprising (II); and (3) a method (IV) of controlling
 CC and/or eradicating a wildtype pestivirus infection. (III) may be used
 CC to protect against wildtype pestivirus infection caused by classical

CC swine fever virus (CSFV), bovine viral diarrhoea virus (BVDV) and Border
 CC disease virus (BDV). When (III) contains Flc7 or Flc9 it may be used as
 CC an emergency vaccination to control outbreaks of CSF when rapid immunity
 CC is required and prophylactic vaccination in areas where classical swine
 CC fever (CSF) is endemic. The present sequence represents a CSFV E-RNS
 CC amino acid sequence given in an example from the present invention.

XX Sequence 228 AA;
 SQ Query Match 60.0%; Score 39; DB 21; Length 228;
 Best Local Similarity 70.0%; Pred. No. 14;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 GVQRAQFQSG 13
 Db 15 GIQRAMFQRG 24
 :||| |||

RESULT 10
 AAW93195
 ID AAW93195 standard; protein; 271 AA.
 XX
 AC AAW93195;
 XX 08-JUN-1999 (first entry)
 DT
 DE Pestivirus Erns cytotoxic polypeptide fragment pErns-NAD.
 XX
 KW Cytotoxic; Erns; ribonuclease activity; pharmaceutical composition;
 KW therapy; medicament; lymphocyte; rheumatism; multiple sclerosis;
 KW organ transplantation; treatment; side effect; anti-fertility;
 KW ribosome-inactivating.
 XX
 OS Pestivirus.
 XX
 PN WO9906569-A1.
 XX
 PD 11-FEB-1999.
 XX
 PF 29-JUL-1998; 98WO-NL00434.
 XX
 PR 30-JUL-1997; 97EP-0202371.
 XX
 PA (DIER-) STICHTING INST DIERHOUDERIJ EN DIERGEZON.
 XX
 PI Bruschke CJM, Van Oirschot JT, Van Rijn PA;
 XX
 DR WPI; 1999-153798/13.
 XX
 PT New selective cytotoxic polypeptide lacking ribonuclease activity -
 PT useful for therapy and for producing medicaments with a selective
 PT cytotoxic effect on lymphocytes
 XX
 PS Disclosure; Fig 4; 41pp; English.
 XX
 CC This invention describes the identification of novel selective cytotoxic
 CC polypeptides (or fragments) partially or wholly lacking ribonuclease
 CC activity. Such polypeptides are derived from a protein having
 CC ribonuclease activity. The polypeptides form a pharmaceutical
 CC composition, which is useful in therapy, and for the production of a
 CC medicament with a selective cytotoxic effect on lymphocytes. Conditions
 CC treated include rheumatism, multiple sclerosis and post organ
 CC transplantation treatment. The use of the selective agents lacking
 CC ribonuclease activity avoids cytotoxic side effects, including
 CC anti-fertility and ribosome-inactivating effects.

XX Sequence 271 AA;
 SQ Query Match 60.0%; Score 39; DB 20; Length 271;
 Best Local Similarity 70.0%; Pred. No. 16;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 GVQRAQFQSG 13

Db 38 GIQRAMFQSG 47
|:|||||

RESULT 11

AAW93196
ID AAW93196 standard; protein; 271 AA.

XX AC
XX AAW93196;

DT 08-JUN-1999 (first entry)

XX Pestivirus Erns cytotoxic polypeptide fragment pErns-sd1.

XX Cytotoxic; Erns; ribonuclease activity; pharmaceutical composition;
KW therapy; medicament; lymphocyte; rheumatism; multiple sclerosis;
KW organ transplantation; treatment; side effect; anti-fertility;
KW ribosome-inactivating.

XX Pestivirus.

XX WO9906569-A1.

XX 11-FEB-1999.

XX 29-JUL-1998; 98WO-NL00434.

XX 30-JUL-1997; 97EP-0202371.

XX (DIER-) STICHTING INST DIERHOUDERIJ EN DIERGEZON.

XX Bruschke CJM, Van Oirschot JT, Van Riijn PA;

XX WPI; 1999-153798/13.

XX New selective cytotoxic polypeptide lacking ribonuclease activity -
PT useful for therapy and for producing medicaments with a selective
PT cytotoxic effect on lymphocytes

XX Disclosure; Fig 4; 4lpp; English.

XX This invention describes the identification of novel selective cytotoxic
CC polypeptides (or fragments) partially or wholly lacking ribonuclease
CC activity. Such polypeptides are derived from a protein having
CC ribonuclease activity. The polypeptides form a pharmaceutical
CC composition, which is useful in therapy, and for the production of a
CC medicament with a selective cytotoxic effect on lymphocytes. Conditions
CC treated include rheumatism, multiple sclerosis and post organ
CC transplantation treatment. The use of the selective agents lacking
CC ribonuclease activity avoids cytotoxic side effects, including
CC anti-fertility and ribosome-inactivating effects.

SQ Sequence 271 AA;

Query Match 60.0%; Score 39; DB 20; Length 271;

Best Local Similarity 70.0%; Pred. No. 16;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 GVQRAQFQSG 13

Db 38 GIQRAMFQSG 47

RESULT 12

AAW93197
ID AAW93197 standard; protein; 271 AA.

XX AC

XX AAW93197;

DT 08-JUN-1999 (first entry)

DE Pestivirus Erns cytotoxic polypeptide fragment pErns-osl.

KW Cytotoxic; Erns; ribonuclease activity; pharmaceutical composition;
KW therapy; medicament; lymphocyte; rheumatism; multiple sclerosis;
KW organ transplantation; treatment; side effect; anti-fertility;
KW ribosome-inactivating.

XX Pestivirus.

XX WO9906569-A1.

XX 11-FEB-1999.

XX 29-JUL-1998; 98WO-NL00434.

XX 30-JUL-1997; 97EP-0202371.

XX (DIER-) STICHTING INST DIERHOUDERIJ EN DIERGEZON.

XX Bruschke CJM, Van Oirschot JT, Van Riijn PA;

XX WPI; 1999-153798/13.

XX New selective cytotoxic polypeptide lacking ribonuclease activity -
PT useful for therapy and for producing medicaments with a selective
PT cytotoxic effect on lymphocytes

XX Disclosure; Fig 4; 4lpp; English.

XX This invention describes the identification of novel selective cytotoxic
CC polypeptides (or fragments) partially or wholly lacking ribonuclease
CC activity. Such polypeptides are derived from a protein having
CC ribonuclease activity. The polypeptides form a pharmaceutical
CC composition, which is useful in therapy, and for the production of a
CC medicament with a selective cytotoxic effect on lymphocytes. Conditions
CC treated include rheumatism, multiple sclerosis and post organ
CC transplantation treatment. The use of the selective agents lacking
CC ribonuclease activity avoids cytotoxic side effects, including
CC anti-fertility and ribosome-inactivating effects.

SQ Sequence 271 AA;

Query Match 60.0%; Score 39; DB 20; Length 271;

Best Local Similarity 70.0%; Pred. No. 16;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 GVQRAQFQSG 13

Db 38 GIQRAMFQSG 47

RESULT 13

AAW93197

ID AAW93197 standard; protein; 453 AA.

XX AC AAW93197;

XX 25-MAR-2003 (updated)

DT 27-APR-1993 (first entry)

DE BDV structural glycoprotein gE1.

XX Bovine Viral Diarrhoea Virus.

XX Bovine Viral Diarrhoea Virus.

XX EP518757-A1.

XX 16-DEC-1992.

XX 10-JUN-1992; 92EP-0401601.

XX 11-JUN-1991; 91FR-0007077.

PA (INMR) RHONE MERIEUX SA.
 PA (EUBI-) SOC EURO BIOTECHNOLOGIE DROIT BELGE SA.
 XX Chappuis G, Pin J, Lecomte CMTG, Renard AJJ;
 XX WPI; 1992-417606/51.
 DR N-PSDB; AAQ32251.

XX New nucleic acid sequences encoding bovine diarrhoea virus
 PT glycoprotein(s) - and derived, recombinant viruses, expression
 PT vectors, proteins etc., useful in protective vaccines
 XX Claim 13; Page 21-22; 37pp; French.

XX This sequence is the gE1 structural protein from the Osloss
 CC strain of BDV (gE1 is called gp62 by Collett et al., Virology 165;
 CC 200-208, 1988). The epitopes of neutralising antibodies which
 CC recognise BDV have been localised to the viral structural proteins.
 CC The invention includes sequences with homology to the nucleotide
 CC sequence encoding gE1, at least the first 273 amino acids of gE2
 CC (also called gp53) and combinations of gE1 with at least part of
 CC gE2 and opt. the other structural (non-glycosylated) protein C
 CC (also called p20) and peptides encoded by this nucleotide
 CC sequences. See also AAQ32249-Q32255.
 CC (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 453 AA;

Query Match 60.0%; Score 39; DB 13; Length 453;
 Best Local Similarity 70.0%; Pred. No. 28;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 GVQRAQFQSG 13
 | : | | | | |
 DB 39 GIQRAMFQSG 48

RESULT 14
 AAR29512
 ID AAR29512 standard; Protein; 731 AA.

XX AAR29512;

XX 25-MAR-2003 (updated)
 DT 27-APR-1993 (first entry)

XX BDV structural glycoprotein gE1 fused to first part of gE2.

XX Bovine Viral Diarrhoea Virus; Osloss; gp53; Pestivirus; Togaviridae;
 XX vaccine; gp62.

XX Bovine Viral Diarrhoea Virus.

XX Key Location/Qualifiers
 FH Region 2..453
 FT /note= "gE1 sequence of AAR29511"
 FT Region 430..727
 FT /note= "gE2 sequence of AAR29509"

XX EP518757-A1.

XX 16-DEC-1992.

XX 10-JUN-1992; 92EP-0401601.

XX 11-JUN-1991; 91FR-0007077.

XX (INMR) RHONE MERIEUX SA.

PA (EUBI-) SOC EURO BIOTECHNOLOGIE DROIT BELGE SA.

XX Chappuis G, Pin J, Lecomte CMTG, Renard AJJ;

XX WPI; 1992-417606/51.

DR N-PSDB; AAQ32252.
 XX New nucleic acid sequences encoding bovine diarrhoea virus
 PT glycoprotein(s) - and derived, recombinant viruses, expression
 PT vectors, proteins etc., useful in protective vaccines
 XX Claim 13; Page 23-25; 37pp; French.

XX This sequence contains a translation initiation methionine, gE1
 CC and the first 273 amino acids of gE2. The structural glycoprotein
 CC coding sequences are from the Osloss strain of BDV (gE1 is called
 CC gp62 and gE2 is called gp53 by Collett et al., Virology 165; 200-208,
 CC 1988). The epitopes of neutralising antibodies which recognise BDV
 CC have been localised to the viral structural proteins. The invention
 CC includes sequences with homology to the nucleotide sequence encoding
 CC at least the first 273 amino acids of gE2, of gE1 alone and of the
 CC combination of gE1 and (part of) gE2 with the other structural protein
 CC C (also called p20) and proteins encoded by these nucleotide sequences.
 CC See also AAQ32249-Q32255.
 CC (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 731 AA;

Query Match 60.0%; Score 39; DB 13; Length 731;
 Best Local Similarity 70.0%; Pred. No. 47;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 GVQRAQFQSG 13
 | : | | | | |
 DB 39 GIQRAMFQSG 48

RESULT 15
 AAR29510
 ID AAR29510 standard; Protein; 823 AA.

XX AAR29510;

XX 25-MAR-2003 (updated)

DT 27-APR-1993 (first entry)

XX BDV structural glycoproteins gE1 and gE2.

XX Bovine Viral Diarrhoea Virus; Osloss; gp53; Pestivirus; Togaviridae;
 XX vaccine; gp62.

XX Bovine Viral Diarrhoea Virus.

XX Key Location/Qualifiers
 FH Region 2..453
 FT /note= "gE1 sequence of AAR29511"
 FT Region 430..727
 FT /note= "gE2 sequence of AAR29509"

XX EP518757-A1.

XX 16-DEC-1992.

XX 10-JUN-1992; 92EP-0401601.

XX 11-JUN-1991; 91FR-0007077.

XX (INMR) RHONE MERIEUX SA.

PA (EUBI-) SOC EURO BIOTECHNOLOGIE DROIT BELGE SA.

XX Chappuis G, Pin J, Lecomte CMTG, Renard AJJ;

XX WPI; 1992-417606/51.

DR N-PSDB; AAQ32250.

XX New nucleic acid sequences encoding bovine diarrhoea virus
 PT glycoprotein(s) - and derived, recombinant viruses, expression
 PT vectors, proteins etc., useful in protective vaccines

```

XX Claim 13; Page 18-20; 37pp; French.
PS
XX
XX This sequence contains a translation initiation methionine, gE1
CC and gE2. The structural glycoprotein coding sequences are from the
CC Ostia strain of BDV (gE1 is called gp62 and gE2 is called gp53
CC by Collett et al., Virology 165: 200-208, 1988). The epitopes of
CC neutralising antibodies which recognise BDV have been localised
CC to the viral structural proteins. The invention includes sequences
CC with homology to the nucleotide sequence encoding at least the
CC first 273 amino acids of gE2, of gE1 alone and of the combination
CC of gE1 and gE2 with the other structural protein C (also called
CC p20) and proteins encoded by these nucleotide sequences.
CC See also AAQ32243-Q32255.
XX (Updated on 25-MAR-2003 to correct PN field.)
SQ Sequence 823 AA;

Query Match 60.0%; Score 39; DB 13; Length 823;
Best Local Similarity 70.0%; Pred. No. 53;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 4 GVORAQFOSG 13
Db 39 GIORAMFORG 48
|:|:|:|:|
|:|:|:|:|

Search completed: January 20, 2004, 16:53:04
Job time : 49.8244 secs

```

1	39	60.0	118	4	US-09-134-001C-5272	Sequence 5272, Appl
2	39	60.0	1286	6	5206163-1	Patent No. 5206163
3	37	56.9	394	4	US-09-252-991A-24485	Sequence 24485, A
4	36.5	56.2	297	4	US-09-252-991A-27478	Sequence 27478, A
5	36.5	56.2	1134	4	US-09-002-285-76	Sequence 76, Appl
6	36.5	56.2	1134	4	US-09-589-477-76	Sequence 76, Appl
7	36	55.4	197	4	US-09-252-991A-18325	Sequence 18325, A
8	36	55.4	437	4	US-09-252-991A-24355	Sequence 24355, A
9	36	55.4	443	4	US-09-252-991A-31911	Sequence 31911, A
10	35	53.8	18	1	US-08-196-945-7	Sequence 7, Appl
11	35	53.8	18	1	US-07-965-667A-8	Sequence 8, Appl
12	35	53.8	18	3	US-08-484-819-8	Sequence 8, Appl
13	35	53.8	18	5	PCT-US93-1C197-8	Sequence 8, Appl
14	35	53.8	42	1	US-08-240-049B-11	Sequence 11, Appl
15	35	53.8	42	3	US-08-478-507-20	Sequence 20, Appl
16	35	53.8	42	3	US-09-123-275A-20	Sequence 20, Appl
17	35	53.8	42	4	US-09-553-427-20	Sequence 20, Appl
18	35	53.8	48	1	US-08-259-148A-13	Sequence 13, Appl
19	35	53.8	48	1	US-08-484-054-13	Sequence 13, Appl
20	35	53.8	48	2	US-07-876-941A-13	Sequence 13, Appl
21	35	53.8	48	3	US-08-542-634-23	Sequence 23, Appl
22	35	53.8	48	3	US-08-477-292-23	Sequence 23, Appl
23	35	53.8	48	4	US-07-870-985A-13	Sequence 13, Appl
24	35	53.8	48	5	PCT-US95-13703-23	Sequence 23, Appl
25	35	53.8	176	1	US-08-196-945-2	Sequence 2, Appl
26	35	53.8	267	4	US-09-172-639-18	Sequence 18, Appl
27	35	53.8	327	1	US-08-240-049B-13	Sequence 13, Appl

```
Query Match      60.0%; Score 39; DB 6; Length 1286;
Best Local Similarity 70.0%; Pred. No. 32;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      4 GVORAFQSG 13
DB      285 GIORAFQSG 294

RESULT 3
US-09-252-991A-24485
; Sequence 24485, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Schnepf, H. Ernest
; APPLICANT: Wicker, Carol
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24485
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24485

Query Match      56.9%; Score 37; DB 4; Length 394;
Best Local Similarity 53.8%; Pred. No. 21;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY      1 APLGVORAFQSG 13
DB      132 ADGLARAFYQSG 144

RESULT 4
US-09-252-991A-27478
; Sequence 27478, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27478
; LENGTH: 297
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27478

Query Match      56.2%; Score 36.5; DB 4; Length 297;
Best Local Similarity 55.6%; Pred. No. 19;
Matches 10; Conservative 0; Mismatches 1; Indels 7; Gaps 1;

QY      3 LGVQ-----RAQFQSG 13
DB      52 LGVQQRFAARAFQFQSG 69

RESULT 5
US-09-002-285-76
; Sequence 76, Application US/09002285
; Patent No. 6369213
; GENERAL INFORMATION:
; APPLICANT: Schnepf, H. Ernest
; APPLICANT: Wicker, Carol
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Walz, Michelle
; APPLICANT: Stockhoff, Brian
; APPLICANT: Muller-Cohn, Judy
; TITLE OF INVENTION: Toxins Active Against Pests
; NUMBER OF SEQUENCES: 105
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/002,285
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/886,615
; FILING DATE: 1-JUL-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/674,002
; FILING DATE: 1-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sanders, Jay W.
; REGISTRATION NUMBER: 39,355
; REFERENCE/DOCKET NUMBER: MA-701C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (352) 375-8100
; TELEFAX: (352) 372-5800
; INFORMATION FOR SEQ ID NO: 76:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1134 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-002-285-76

Query Match      56.2%; Score 36.5; DB 4; Length 1134;
Best Local Similarity 64.3%; Pred. No. 86;
Matches 9; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY      1 APL-GVORAFQSG 13
DB      400 APLFGITRAQFYPG 413

RESULT 6
US-09-589-477-76
; Sequence 76, Application US/09589477
; Patent No. 6570005
; GENERAL INFORMATION:
; APPLICANT: Schnepf, H. Ernest
; APPLICANT: Wicker, Carol
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Walz, Michelle
; APPLICANT: Stockhoff, Brian
; TITLE OF INVENTION: Toxins Active Against Pests
; NUMBER OF SEQUENCES: 105
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
```

CITY: Gainesville
STATE: Florida
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/589,477
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/886,615
FILING DATE: 1-JUL-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/674,002
FILING DATE: 1-JUL-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Sanders, Jay M.
REGISTRATION NUMBER: 39,355
REFERENCE/DOCKET NUMBER: MA-701C1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (352) 375-8100
TELEFAX: (352) 372-5800
INFORMATION FOR SEQ ID NO: 76:
SEQUENCE CHARACTERISTICS:
LENGTH: 1134 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-589-477-76

```

Query Match      56.2%; Score 36.5; DB 4; Length 1134;
Best Local Similarity 64.3%; Pred. No. 86;
Matches 9; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY      1 APL-GVORAQFQSG 13
      ||| | | | | |
      400 APLFGITRAQFYFG 413

Db

RESULT 7
US-09-252-991A-18325
; Sequence 18325, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18325
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18325

Query Match      55.4%; Score 36; DB 4; Length 197;
Best Local Similarity 63.8%; Pred. No. 15;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      3 LGVORAQFQSG 13

```

Db 81 LGARGEVFSG 91

RESULT 8

US-09-252-991A-24355

; Sequence 24355, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 24355

; LENGTH: 437

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-24355

Query Match 55.4%; Score 36; DB 4; Length 437;

Best Local Similarity 58.3%; Pred. No. 36;

Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 2 PLGVQRAQFQSG 13

Db 331 PLGDARAEELQNG 342

RESULT 9

US-09-252-991A-31911

; Sequence 31911, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 31911

; LENGTH: 443

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-31911

Query Match 55.4%; Score 36; DB 4; Length 443;

Best Local Similarity 70.0%; Pred. No. 37;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 APLGVQRAQF 10

Db 218 APASVERAQF 227

RESULT 10

US-08-196-945-7

; Sequence 7, Application US/08196945

; Patent No. 5563032

; GENERAL INFORMATION:

; APPLICANT: KHUDYAKOV, YURI E.

; APPLICANT: FIELDS, HOWARD A.


```

; APPLICANT: FAVOROV, MICHAEL O.
; TITLE OF INVENTION: MOSAIC POLYPEPTIDE AND METHODS FOR
; TITLE OF INVENTION: DETECTING THE HEPATITIS E VIRUS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: Suite 1200, The Candler Bldg.,
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/196,945
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.039
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-196-945-7

Query Match 53.8%; Score 35; DB 1; Length 18;
Best Local Similarity 63.6%; Pred. No. 1.5;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 PLGVQRAQFQS 12
Db 2 PLGLQGCAFQS 12

RESULT 11
US-07-965-667A-8
; Sequence 8, Application US/07965667A
; Patent No. 5736315
; GENERAL INFORMATION:
; APPLICANT: Fields, Howard A.
; APPLICANT: Favorov, Michael O.
; APPLICANT: Khudiyakov, Yuri E.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTING
; TITLE OF INVENTION: ANTI-HEPATITIS E VIRUS ACTIVITY
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 133 Carnegie Way, N.W., Suite 400
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/965,667A
; FILING DATE:
; CLASSIFICATION: 514

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.022
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 688-0770
; TELEFAX: (404) 688-9880
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-07-965-667A-8

Query Match 53.8%; Score 35; DB 1; Length 18;
Best Local Similarity 63.6%; Pred. No. 1.5;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 PLGVQRAQFQS 12
Db 2 PLGLQGCAFQS 12

RESULT 12
US-08-484-819-8
; Sequence 8, Application US/08484819
; Patent No. 6022685
; GENERAL INFORMATION:
; APPLICANT: Fields, Howard A.
; APPLICANT: Favorov, Michael O.
; APPLICANT: Khudiyakov, Yuri E.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTING
; TITLE OF INVENTION: ANTI-HEPATITIS E VIRUS ACTIVITY
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 133 Carnegie Way, N.W., Suite 400
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,819
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/965,667
; FILING DATE: 21-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.022
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 688-0770
; TELEFAX: (404) 688-9880
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-484-819-8

Query Match 53.8%; Score 35; DB 3; Length 18;
Best Local Similarity 63.6%; Pred. No. 1.5;

```

Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 PLGVQRAQFQS 12
|||:|
Db 2 PLGLQGCAQFQS 12

RESULT 13
PCT-US93-10197-8
; Sequence 8, Application PC/TUS9310197
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: "METHODS AND COMPOSITIONS FOR DETECTING
; TITLE OF INVENTION: ANTI-HEPATITIS E VIRUS ACTIVITY"
; NUMBER OF SEQUENCES: 20
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 EPO
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/10197
; PRIOR APPLICATION DATA:
; FILING DATE: 21-OCT-1992
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
PCT-US93-10197-8

Query Match 53.8%; Score 35; DB 5; Length 18;
Best Local Similarity 63.6%; Pred. No. 1.5;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 PLGVQRAQFQS 12
|||:|
Db 2 PLGLQGCAQFQS 12

RESULT 14
US-08-240-049B-11
; Sequence 11, Application US/08240049B
; Patent No. 5686239
; GENERAL INFORMATION:
; APPLICANT: Reyes, Gregory R.
; APPLICANT: Tam, Albert W.
; APPLICANT: Varbough, Patrice O.
; TITLE OF INVENTION: Hepatitis E Virus Peptides and Methods
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genelabs Technologies, Inc.
; STREET: 505 Penobscot Drive
; CITY: Redwood City
; STATE: CA
; COUNTRY: USA
; ZIP: 94063
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/240,049B
; FILING DATE: 09-MAY-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles K. Sholtz
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0093.35; G32P4

TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Hepatitis E Virus (Burma strain)
; INDIVIDUAL ISOLATE: 406.3-2
US-08-240-049B-11

Query Match 53.8%; Score 35; DB 1; Length 42;
Best Local Similarity 63.6%; Pred. No. 4;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 PLGVQRAQFQS 12
|||:|
Db 20 PLGLQGCAQFQS 30

RESULT 15
US-08-478-507-20
; Sequence 20, Application US/08478507
; Patent No. 6120988
; GENERAL INFORMATION:
; APPLICANT: Reyes, Gregory R.
; APPLICANT: Varbough, Patrice O.
; APPLICANT: Bradley, Daniel W.
; APPLICANT: Krawczynski, Krzysztof Z.
; APPLICANT: Tam, Albert
; APPLICANT: Fry, Kirk E.
; TITLE OF INVENTION: DNA Sequences of Enterically Transmitted
; TITLE OF INVENTION: No. 6120988-A/No. 6120988-B Hepatitis Viral Agent
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,507
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/279,823
; FILING DATE: 25-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/681,078
; FILING DATE: 05-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/505,888
; FILING DATE: 05-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/420,921
; FILING DATE: 13-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/367,486
; FILING DATE: 16-JUN-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/336,672
; FILING DATE: 11-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/208,997

FILING DATE: 17-JUN-1988
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0183.22
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 324-0880
TELEFAX: (650) 324-0960
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 42 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: 406.3-2 epitope - Burma strain
US-08-478-507-20

Query Match 53.8%; Score 35; DB 3; Length 42;
Best Local Similarity 63.6%; Pred. No. 4;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 PLGVQRAQFQS 12
|||:|
Db 20 PLGLQGCAQFS 30

Search completed: January 20, 2004, 17:47:14
Job time : 16.2824 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:54:27 ; Search time 34.4351 Seconds
(without alignments)
77.196 Million cell updates/sec

Title: US-10-089-549-18
Perfect score: 65
Sequence: 1 APLGVQRAQFQSG 13

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 762491 seqs, 204481190 residues

Total number of hits satisfying chosen parameters: 762491

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*

```
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY-ES

Result No.	Score	Query Match	Length	ID	Description
1	65	100.0	13	12	US-10-089-549-18
2	38	58.5	818	15	US-10-026-994-2
3	37	56.9	227	16	US-10-312-062-6
4	37	56.9	897	16	US-10-312-062-2
5	36.5	56.2	1134	15	US-10-099-285-76
6	36	55.4	936	12	US-10-017-161-1676
7	36	55.4	936	12	US-10-292-798-1336
8	35	53.8	42	11	US-09-851-410-20
9	35	53.8	48	10	US-09-769-066-23
10	35	53.8	48	11	US-09-468-147-172
11	35	53.8	48	12	US-10-165-868-13
12	35	53.8	280	15	US-10-183-702A-4
13	35	53.8	327	10	US-09-769-066-17
14	35	53.8	327	12	US-10-165-868-15
15	35	53.8	436	12	US-10-185-868-17

16	53.8	459	15	US-10-078-167-3	Sequence 3, Appli
17	53.8	473	12	US-10-369-493-3972	Sequence 3972, Ap
18	53.8	529	12	US-10-369-493-7248	Sequence 7248, Ap
19	53.8	530	16	US-10-314-739-3	Sequence 3, Appli
20	53.8	537	12	US-10-369-493-4489	Sequence 4489, Ap
21	53.8	540	10	US-09-769-066-25	Sequence 25, Appl
22	53.8	549	10	US-09-769-066-15	Sequence 15, Appl
23	53.8	660	10	US-09-769-066-13	Sequence 13, Appl
24	53.8	660	11	US-09-851-410-8	Sequence 8, Appli
25	53.8	660	12	US-10-165-868-19	Sequence 19, Appl
26	53.8	660	12	US-10-257-044-1	Sequence 1, Appli
27	52.3	52	15	US-10-106-698-5182	Sequence 5182, Ap
28	52.3	139	9	US-09-864-761-40558	Sequence 40558, A
29	52.3	142	15	US-10-106-698-5042	Sequence 5042, Ap
30	52.3	174	12	US-10-017-161-1556	Sequence 1556, Ap
31	52.3	234	15	US-10-156-761-13779	Sequence 13779, A
32	52.3	282	12	US-10-369-493-15356	Sequence 15356, A
33	52.3	289	12	US-10-369-493-15723	Sequence 15723, A
34	52.3	289	12	US-10-369-493-16108	Sequence 16108, A
35	52.3	605	15	US-10-190-555-2	Sequence 2, Appli
36	52.3	61	12	US-10-029-386-30183	Sequence 30183, A
37	50.8	99	9	US-09-764-860-305	Sequence 305, App
38	50.8	99	12	US-10-212-872-305	Sequence 305, App
39	50.8	99	15	US-10-074-095-305	Sequence 98, Appl
40	50.8	129	11	US-09-992-600A-98	Sequence 98, Appl
41	50.8	129	11	US-09-924-340-98	Sequence 98, Appl
42	50.8	129	12	US-09-992-095B-98	Sequence 98, Appl
43	50.8	129	12	US-10-154-678-98	Sequence 98, Appl
44	50.8	129	12	US-09-999-570-98	Sequence 98, Appl
45	50.8	129	15	US-10-000-489-98	Sequence 98, Appl

ALIGNMENTS

RESULT 1
US-10-089-549-18
; Sequence 18, Application US/10089549
; Publication No. US20030194762A1
; GENERAL INFORMATION:
; APPLICANT: KUBOTA, Michio
; APPLICANT: TSUSAKI, Keiji
; APPLICANT: HIGASHIYAMA, Takanobu
; APPLICANT: FUKUDA, Shigeharu
; APPLICANT: MIYAKE, Toshio
; TITLE OF INVENTION: ALPHA-ISOMALTOSYLGLUCOSACCHARIDE-FORMING ENZYME, PROCESS AND U
; TITLE OF INVENTION: SAME
; FILE REFERENCE: KUBOTA-9
; CURRENT APPLICATION NUMBER: US/10/089,549
; CURRENT FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: JP 233364/2000
; PRIOR FILING DATE: 2000-08-01
; PRIOR APPLICATION NUMBER: JP 234937/2000
; PRIOR FILING DATE: 2000-08-02
; PRIOR APPLICATION NUMBER: PCT/JP01/06412
; PRIOR FILING DATE: 2001-07-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 18
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Arthrobacter globiformis
US-10-089-549-18

Query Match 100.0%; Score 65; DB 12; Length 13;
Best Local Similarity 100.0%; Pred.No. 4.7e-06;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 APLGVQRAQFQSG 13
Db 1 APLGVQRAQFQSG 13

```
RESULT 2
US-10-026-994-2
; Sequence 2, Application US/10026994
; Publication No. US20030113732A1
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Goedegebuur, Frits
; APPLICANT: Ward, Michael
; APPLICANT: Yao, Jian
; TITLE OF INVENTION: EGY Endoglucanase and Nucleic Acids
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: GC698
; CURRENT APPLICATION NUMBER: US/10/026,994
; CURRENT FILING DATE: 2002-04-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 818
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-10-026-994-2
Query Match 58.5%; Score 38; DB 15; Length 818;
Best Local Similarity 72.7%; Pred. No. 90;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 APLGVQRAQFQ 11
Db 541 ASSGVQSQFQ 551

RESULT 3
US-10-312-062-6
; Sequence 6, Application US/10312062
; Publication No. US20030129744A1
; GENERAL INFORMATION:
; APPLICANT: Schlapp, Tobias
; APPLICANT: Rijsewijk, Franciscus A. M.
; TITLE OF INVENTION: BVD Virus-Like Particles
; FILE REFERENCE: Le A 34 440
; CURRENT APPLICATION NUMBER: US/10/312,062
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: PCT/EP01/06738
; PRIOR FILING DATE: 2001-06-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
; LENGTH: 227
; TYPE: PRT
; ORGANISM: Bovine Viral Diarrhea Virus
US-10-312-062-6
Query Match 56.9%; Score 37; DB 16; Length 227;
Best Local Similarity 70.0%; Pred. No. 34;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 GVQRAQFQSG 13
Db 15 GVQRAFMFEG 24

RESULT 4
US-10-312-062-2
; Sequence 2, Application US/10312062
; Publication No. US20030129744A1
; GENERAL INFORMATION:
; APPLICANT: Schlapp, Tobias
; APPLICANT: Rijsewijk, Franciscus A. M.
; TITLE OF INVENTION: BVD Virus-Like Particles
; FILE REFERENCE: Le A 34 440
; CURRENT APPLICATION NUMBER: US/10/312,062
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: PCT/EP01/06738
US-10-099-285-76
; Sequence 76, Application US/10099285
; Publication No. US20030105319A1
; GENERAL INFORMATION:
; APPLICANT: Schnepf, H. Ernest
; APPLICANT: Wicker, Carol
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Walz, Michelle
; APPLICANT: Stockhoff, Brian
; APPLICANT: Muller-Cohn, Judy
; TITLE OF INVENTION: Toxins Active Against Pests
; NUMBER OF SEQUENCES: 105
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/099,285
; FILING DATE: 15-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/002,285
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US 08/886,615
; FILING DATE: 1-JUL-1997
; APPLICATION NUMBER: US 08/674,002
; FILING DATE: 1-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sanders, Jay M.
; REGISTRATION NUMBER: 39,355
; REFERENCE/DOCKET NUMBER: MA-701C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (352) 375-8100
; TELEFAX: (352) 372-5800
; INFORMATION FOR SEQ ID NO: 76:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1134 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 76:
US-10-099-285-76
Query Match 56.2%; Score 36.5; DB 15; Length 1134;
Best Local Similarity 64.3%; Pred. No. 2.5e+02;
```

Qy 1 APLGVQRAQ 9
835 APLGLOREO 843
Db

L-410-20

QY 2 PLGVQRAQFQS 12
|||:|
Db 20 PLGLOGCAFOS 30

RESULT 9

US-09-769-066-23

Sequence 23, Application US/09763066

Patent No. US20020107360A1

GENERAL INFORMATION:

APPLICANT: Fuerst, Thomas R.

McAtee, C. Patrick

Yarbough, Patrice O.

Zhang, Yifan

TITLE OF INVENTION: HEPATITIS E VIRUS ANTIGENS AND USES THEREFOR

NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dehlinger & Associates

STREET: 350 Cambridge Ave., Suite 250

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/769,066

FILING DATE: 24-Jan-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/542,634

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Fabian, Gary R.

REGISTRATION NUMBER: 33,875

REFERENCE/DOCKET NUMBER: 4600-0293.30

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 324-0880

TELEFAX: (415) 324-0960

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:

LENGTH: 48 amino acids

TYPE: amino acid

STRANDEDNESS: Hepatitis E Virus (Burma strain)

406.3-2

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ORIGINAL SOURCE:

SEQUENCE DESCRIPTION: SEQ ID NO: 23:

US-09-769-066-23

Query Match 53.8%; Score 35; DB 10; Length 48;

Best Local Similarity 63.6%; Pred. No. 15;

Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 PLGVORAFQFS 12

DB 20 PLGLQGCAFS 30

RESULT 10

US-09-468-147-172

Sequence 172, Application US/09468147A

Publication No. US20030049601A1

GENERAL INFORMATION:

APPLICANT: Abbott Laboratories

APPLICANT: Schlauder, George G.

APPLICANT: Erker, James C.

APPLICANT: Desai, Suresh M.

APPLICANT: Dawson, George J.

APPLICANT: Mushahwar, I. K.

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTING

HEPATITIS E VIRUS

FILE REFERENCE: 6232.US.P1

CURRENT APPLICATION NUMBER: US/09/468,147A

CURRENT FILING DATE: 1999-12-21

EARLIER APPLICATION NUMBER: US 09/173,141

EARLIER FILING DATE: 1998-10-15

EARLIER APPLICATION NUMBER: US 60/061,199

EARLIER FILING DATE: 1997-10-15

NUMBER OF SEQ ID NOS: 258

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 172

LENGTH: 48

TYPE: PRT

ORGANISM: Hepatitis E Virus

FEATURE:

OTHER INFORMATION: B 3-2e

US-09-468-147-172

Query Match 53.8%; Score 35; DB 11; Length 48;

Best Local Similarity 63.6%; Pred. No. 15;

Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 PLGVORAFQFS 12

DB 20 PLGLQGCAFS 30

RESULT 11

US-10-165-868-13

Sequence 13, Application US/10165868

Publication No. US20030143241A1

GENERAL INFORMATION:

APPLICANT: Reyes, Gregory R.

Bradley, Daniel W.

Tsu, Jr-Shin

Purdy, Michael A.

Tam, Albert W.

Krawczynski, Krzysztof Z.

TITLE OF INVENTION: Hepatitis E Virus Vaccine and Method

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dehlinger & Associates

STREET: 350 Cambridge Avenue, Suite 250

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/165,868

FILING DATE: 06-Jun-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/07/870,985A

FILING DATE: 20-APRIL-1992

APPLICATION NUMBER: US 822,335

FILING DATE: 17-JAN-1992

APPLICATION NUMBER: US 505,888

FILING DATE: 05-APRIL-1990

APPLICATION NUMBER: US 420,921

FILING DATE: 13-OCTOBER-1989

APPLICATION NUMBER: US 367,486

FILING DATE: 16-JUNE-1989

APPLICATION NUMBER: US 336,672

FILING DATE: 11-APRIL-1989

APPLICATION NUMBER: US 208,997

FILING DATE: 17-JUNE-1988

ATTORNEY/AGENT INFORMATION:

NAME: Fabian, Gary R.

REGISTRATION NUMBER: 33,875

REFERENCE/DOCKET NUMBER: 4600-0093.30

Wed Jan 21 11:28:38 2004

```
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 48 amino acids
TYPE: amino acid
MOLECULE TYPE: linear
TOPOLOGY: linear
HYPOTHETICAL: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: 406.3-2, BURMA, FIGURE 9
SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-10-165-868-13

Query Match 53.8%; Score 35; DB 12; Length 48;
Best Local Similarity 63.6%; Pred. No. 15;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 PLGVQRAQFQS 12
|||:|
Db 20 PLGLQCAQFS 30

RESULT 12
US-10-188-702A-4
Sequence 4, Application US/10188702A
Publication No. US20030110533A1
GENERAL INFORMATION:
APPLICANT: Rebecca E. Cahoon
APPLICANT: Leslie T. Harvell
APPLICANT: Anthony J. Kinney
APPLICANT: Yong Tao
TITLE OF INVENTION: Acetyl-CoA Carboxylase Subunits
FILE REFERENCE: BE1480USNA
CURRENT APPLICATION NUMBER: US/10/188,702A
CURRENT FILING DATE: 2002-07-03
PRIOR APPLICATION NUMBER: 60/303,387
PRIOR FILING DATE: 2001-07-06
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Microsoft Office 97
SEQ ID NO 4
LENGTH: 280
TYPE: PRT
ORGANISM: Glycine max
US-10-188-702A-4

Query Match 53.8%; Score 35; DB 15; Length 280;
Best Local Similarity 57.1%; Pred. No. 1.1e+02;
Matches 8; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

QY 2 PLGV--QRAQFQS 13
|||:|
Db 21 PLGLNSQKVSFQS 34

RESULT 13
US-09-769-066-17
Sequence 17, Application US/09769066
Patent No. US20020107360A1
GENERAL INFORMATION:
APPLICANT: Fuerst, Thomas R.
APPLICANT: McAtee, C. Patrick
APPLICANT: Yarbough, Patrice O.
APPLICANT: Zhang, Yifan
TITLE OF INVENTION: HEPATITIS E VIRUS ANTIGENS AND USES THEREFOR
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Ave., Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA

ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/769,066
FILING DATE: 24-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/542,634
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 4600-0293.30
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 327 amino acids
TYPE: amino acid
STRANDEDNESS: Hepatitis E Virus (Burma strain) SG3
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-769-066-17

Query Match 53.8%; Score 35; DB 10; Length 327;
Best Local Similarity 63.6%; Pred. No. 1.3e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 PLGVQRAQFQS 12
|||:|
Db 299 PLGLQCAQFS 309

RESULT 14
US-10-165-868-15
Sequence 15, Application US/10165868
Publication No. US20030143241A1
GENERAL INFORMATION:
APPLICANT: Reyes, Gregory R.
APPLICANT: Bradley, Daniel W.
APPLICANT: Twu, Jr-Shin
APPLICANT: Purdy, Michael A.
APPLICANT: Tam, Albert W.
APPLICANT: Krawczynski, Krzysztof Z.
TITLE OF INVENTION: Hepatitis E Virus Vaccine and Method
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/165,868
FILING DATE: 06-Jun-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/870,985A
```


FILING DATE: 20-APRIL-1992
APPLICATION NUMBER: US 822,335
FILING DATE: 17-JAN-1992
APPLICATION NUMBER: US 505,888
FILING DATE: 05-APRIL-1990
APPLICATION NUMBER: US 420,921
FILING DATE: 13-OCTOBER-1989
APPLICATION NUMBER: US 367,486
FILING DATE: 16-JUNE-1989
APPLICATION NUMBER: US 336,672
FILING DATE: 11-APRIL-1989
APPLICATION NUMBER: US 208,997
FILING DATE: 17-JUNE-1988
ATTORNEY/AGENT INFORMATION:
NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 4600-0093.30
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 327 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: SG3, BURMA, FIGURE 9
SEQUENCE DESCRIPTION: SEQ ID NO: 15:
US-10-165-868-15
Query Match 53.8%; Score 35; DB 12; Length 327;
Best Local Similarity 63.6%; Pred. No. 1.3e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 2 PLGVQRAQFQS 12
DB 299 PLGLQCAQFQS 309
RESULT 15
US-10-165-868-17
Sequence 17, Application US/10165868
Publication No. US20030143241A1
GENERAL INFORMATION:
APPLICANT: Reyes, Gregory R.
Bradley, Daniel W.
Twu, Jr-Shin
Purdy, Michael A.
Tam, Albert W.
Krawczynski, Krzysztof Z.
TITLE OF INVENTION: Hepatitis E Virus Vaccine and Method
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/165,868
FILING DATE: 06-Jun-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/870,985A
FILING DATE: 20-APRIL-1992

APPLICATION NUMBER: US 822,335
FILING DATE: 17-JAN-1992
APPLICATION NUMBER: US 505,888
FILING DATE: 05-APRIL-1990
APPLICATION NUMBER: US 420,921
FILING DATE: 13-OCTOBER-1989
APPLICATION NUMBER: US 367,486
FILING DATE: 16-JUNE-1989
APPLICATION NUMBER: US 336,672
FILING DATE: 11-APRIL-1989
APPLICATION NUMBER: US 208,997
FILING DATE: 17-JUNE-1988
ATTORNEY/AGENT INFORMATION:
NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 4600-0093.30
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 436 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: C2, BURMA, FIGURE 9
SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-10-165-868-17
Query Match 53.8%; Score 35; DB 12; Length 436;
Best Local Similarity 63.6%; Pred. No. 1.8e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 2 PLGVQRAQFQS 12
DB 408 PLGLQCAQFQS 418
Search completed: January 20, 2004, 17:53:17
Job time : 35.4351 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:43:26 ; Search time 13.4962 Seconds
(without alignments)
92.633 Million cell updates/sec

Title: US-10-089-549-18
Perfect score: 65
Sequence: 1 APLGVQRAQFQSG 13

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	60.0	2245	T18278	myosin heavy chain
2	39	60.0	3898	A44217	genome polyprotein
3	39	60.0	3988	GNWVBY	genome polyprotein
4	37	56.9	376	D83099	probable RND efflu
5	36.5	56.2	1154	S39536	parasporal crystal
6	36	55.4	381	T30452	hypothetical prote
7	36	55.4	399	A82351	DNA/pantothenate m
8	36	55.4	417	S16582	fructose-bisphosph
9	36	55.4	417	T47564	fructose-bisphosph
10	36	55.4	2051	T30938	receptor tyrosine
11	35	53.8	192	AC0990	conserved hypochet
12	35	53.8	237	G75476	conserved hypochet
13	35	53.8	268	T31699	hypothetical prote
14	35	53.8	358	S43876	site-specific DNA
15	35	53.8	358	A87296	modification methy
16	35	53.8	482	B87382	hypothetical prote
17	35	53.8	509	T07721	hypothetical prote
18	35	53.8	530	T337471	capsid protein - N
19	35	53.8	660	1VHWI2	structural protein
20	35	53.8	785	S22367	phenylalanine-tRNA
21	35	53.8	785	T52303	phenylalanine-tRNA
22	35	53.8	1978	S77257	hypothetical prote
23	34	52.3	209	S10834	hypothetical prote
24	34	52.3	247	I40228	spoIV protein - Ba
25	34	52.3	317	A81402	probable integral
26	34	52.3	319	F82601	5'-nucleotidase XF
27	34	52.3	346	T08864	hypothetical prote
28	34	52.3	375	C64824	probable RNA methy
29	34	52.3	375	C90746	probable enzyme [1

30	34	52.3	375	2	G85596	probable enzyme yb
31	34	52.3	389	2	E69727	biosynthesis of te
32	34	52.3	395	2	G87481	N-acetylmutamoyl-L
33	34	52.3	399	2	H87342	glycosyl transfera
34	34	52.3	415	2	T40338	ccat-binding fact
35	34	52.3	430	2	T45759	MS5-like protein -
36	34	52.3	433	2	T36122	probable integral
37	34	52.3	454	2	E87571	amine oxidase, fla
38	34	52.3	546	2	D84555	probable protein k
39	34	52.3	693	1	A41090	serine/threonine-s
40	34	52.3	876	2	A45956	beta-glucosidase {
41	34	52.3	938	2	C84480	hypothetical prote
42	34	52.3	1840	2	T30250	GPI protein - mous
43	33	50.8	133	2	AC3949	transcription regu
44	33	50.8	133	2	G98333	probable transcript
45	33	50.8	168	2	AI2865	transcription regu

ALIGNMENTS

RESULT 1

T18278
myosin heavy chain [similarity] - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000
C:Accession: T18278; T30579
R:Hammer III, J.A.; Jung, G.
J. Biol. Chem. 271, 7120-7127, 1996
A:Title: The sequence of the dictyostelium myo J heavy chain gene predicts a novel, d
A:Reference number: Z18954; MUID:96215148; PMID:8636147
A:Accession: T18278
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2245 <HAM>
A:Cross-references: EMBL:U42409; NID:g1150765; PID:g1150766; FIDN:AAA85186.1
R:Titus, M.A.; Kuspa, A.; Loomis, W.F.
Proc. Natl. Acad. Sci. U.S.A. 91, 9446-9450, 1994
A:Title: Discovery of myosin genes by physical mapping in Dictyostelium.
A:Reference number: Z20873; MUID:95023928; PMID:7937787
A:Accession: T30579
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-190, 'F', 192-283, 'T', 285-290, 'R', 292-331, 'IE', 'WW', 337-338, 'LK', 342, 'YRM'
A:Cross-references: EMBL:L35322; NID:g1039360; PID:g1039361; FIDN:AAA79858.1
C:Genetics:
A:Gene: myoJ
A:Introns: 257/3; 307/3
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: ATP; P-loop
F:84-809/Domain: myosin motor domain homology <MMO>

Query Match 60.0%; Score 39; DB 2; Length 2245;
Best Local Similarity 72.7%; Pred. No. 61;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 3 LGVQRAQFQSG 13
||| :|||
Db 1190 LGQSSQFQSG 1200

RESULT 2

A44217
genome polyprotein - bovine viral diarrhea virus (strain SD-1)
C:Species: bovine viral diarrhea virus, BVDV
C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 19-Jan-2001
C:Accession: A44217
R:Deng, R.; Brock, K.V.
Virology 191, 867-879, 1992
A:Title: Molecular cloning and nucleotide sequence of a pestivirus genome, noncytopat
A:Reference number: A44217; MUID:93079889; PMID:1333126
A:Accession: A44217
A:Molecule type: genomic RNA

A:Residues: 1-3898 <DEN>
 A:Cross-references: GB:M96751; NID:Q289507; PIDN:AAA42860.1; PID:G289508
 A:Note: this polyprotein may be cleaved into several mature proteins, including p23 prot and p75 protein; the cleavage sites are not reported
 C:Superfamily: pestivirus genome polyprotein
 C:Keywords: ATP; glycoprotein; nucleotide binding; P-loop; polyprotein; RNA binding; zinc
 F:2-234/Product: viral proteinase p20 #status predicted <VPT>
 F:253-265/Region: hydrophobic
 F:546-1115/Product: major envelope glycoprotein gp55 #status predicted <EGP>
 F:547-562/Region: hydrophobic
 F:556-670/Region: hydrophobic
 F:675-694/Region: hydrophobic
 F:1031-1046/Region: hydrophobic
 F:1074-1099/Region: hydrophobic
 F:1149-1164/Region: hydrophobic
 F:1217-1238/Region: hydrophobic
 F:1252-1269/Region: hydrophobic
 F:1271-1292/Region: hydrophobic
 F:1293-1304/Region: hydrophobic
 F:1357-1373/Region: hydrophobic
 F:1484-1512/Region: zinc finger
 F:1815-1822/Region: nucleotide-binding motif A (P-loop)
 F:1906-1911/Region: nucleotide-binding motif B
 F:1910-1913/Region: DEAH motif
 F:2562-2582/Region: hydrophobic
 F:2722,281,296,335,370,413,487,597,809,922,990,1357,1419,1713,2134,2217,2494,2682

Query Match 60.0%; Score 39; DB 1; Length 3898;
 Best Local Similarity 70.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 GVQRAQFQSG 13
 DB 285 GIQRAMFQSG 294

RESULT 3
 GNWVAV
 genome polyprotein - bovine viral diarrhoea virus
 C:Species: bovine viral diarrhoea virus, BVDV
 C:Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 19-Jan-2001
 R:Collett, M.S.; Larson, R.; Gold, C.; Strick, D.; Anderson, D.K.; Purchio, A.F.
 Virolology 165, 191-199, 1983
 A:Title: Molecular cloning and nucleotide sequence of the pestivirus bovine viral diarrhoea virus
 A:Reference number: A29198; MUID:88265858; PMID:2838957
 A:Accession: A29198
 A:Molecule type: genomic RNA
 A:Residues: 1-3988 <COL>
 A:Experimental source: isolate NADL
 R:Ward, P.; Misra, V.
 Am J. Vet. Res. 52, 1231-1236, 1991
 A:Title: Detection of bovine viral diarrhoea virus, using degenerate oligonucleotide primers
 A:Reference number: A61161; MUID:92027091; PMID:1656820
 A:Accession: A61161
 A:Molecule type: genomic RNA
 A:Residues: 2054-2072 <WAR>
 A:Experimental source: isolate V1352
 A:Note: authors translated the codon ATA for residue 18 as Thr
 C:Superfamily: pestivirus genome polyprotein
 C:Keywords: ATP; glycoprotein; nucleotide binding; P-loop; polyprotein
 F:2-234/Product: viral proteinase p20 #status predicted <VPT>
 F:548-1115/Product: major envelope glycoprotein gp55 #status predicted <EGP>
 F:1905-1912/Region: nucleotide-binding motif A (P-loop)
 F:1996-2001/Region: nucleotide-binding motif B
 F:2000-2003/Region: DEAH motif
 F:2722,281,296,335,370,413,487,597,809,922,990,1357,1419,1451,1803,2224,2307,2584,2772,29

Query Match 60.0%; Score 39; DB 1; Length 3898;
 Best Local Similarity 70.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 GVQRAQFQSG 13

DB 285 GIQRAMFQSG 294

RESULT 4
 DB3099
 probable RND efflux membrane fusion protein precursor PA4374 [imported] - Pseudomonas aeruginosa
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: D83099
 R:Stover, C.K.; Ham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
 A:Reference number: A82950; MUID:20437337; PMID:10984043
 A:Accession: D83099
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-376 <STO>
 A:Cross-references: GB:AE004853; GB:AE004091; NID:G9950598; PIDN:AAG07762.1; GSPDB:G
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA4374

Query Match 56.9%; Score 37; DB 2; Length 376;
 Best Local Similarity 53.8%; Pred. No. 23;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 APLGVORAFQSG 13
 DB 114 ADLGLARAEYQSG 126

RESULT 5
 S39536
 parasporal crystal protein cry98a1 - Bacillus thuringiensis
 N:Alternate names: delta-endotoxin-related protein; parasporal crystal protein cryX
 C:Species: Bacillus thuringiensis
 C:Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 01-Dec-2000
 C:Accession: S39536
 R:Shevelev, A.B.; Svarinskaya, M.A.; Karasin, A.I.; Kogan, Y.N.; Chestukhina, G.G.; St PERS Lett. 336, 79-82, 1993
 A:Title: Primary structure of the cryX, the novel delta-endotoxin-related gene from Bacillus thuringiensis
 A:Reference number: S39536; MUID:94085596; PMID:8262221
 A:Accession: S39536
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1154 <SHE>
 A:Cross-references: EMBL:X75019
 C:Superfamily: parasporal crystal protein
 C:Keywords: delta-endotoxin

Query Match 56.2%; Score 36.5; DB 2; Length 1154;
 Best Local Similarity 64.3%; Pred. No. 94;
 Matches 9; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 1 APLGVORAFQSG 13
 DB 420 APLGLITRAQFYFG 433

RESULT 6
 T30452
 hypothetical protein ORF102 - Lymantria dispar nuclear polyhedrosis virus
 C:Species: Lymantria dispar nuclear polyhedrosis virus, LdNPV
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jul-2000
 C:Accession: T30452
 R:Kuzio, J.; Pearson, M.N.; Harwood, S.H.; Funk, C.J.; Evans, J.T.; Slavicek, J.M.; Virolology 253, 17-34, 1999
 A:Title: Sequence and analysis of the genome of a baculovirus pathogenic for Lymantria dispar
 A:Reference number: Z20836; MUID:99124785; PMID:9887315
 A:Accession: T30452

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-381 <KUZ>

A:Cross-references: EMBL:AF081810; NID:G3822234; PIDN:AACT5288.1; PID:G3822337

Query Match 55.4%; Score 36; DB 2; Length 381;

Best Local Similarity 53.8%; Pred. No. 37;

Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 APLGVQRAQFQSG 13

Db 110 APLGSRRAQTAAG 122

RESULT 7

DBA/pantothenate metabolism flavoprotein VC0215 [imported] - Vibrio cholerae (strain N16)

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C:Accession: A82351

R:Heidelberger, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;

chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.

1, R.R.; Moxallanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833; PMID:10952301

A:Accession: A82351

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-399 <HEZ>

A:Cross-references: GB:AE004111; GB:AE003852; NID:G9654614; PIDN:AAF93391.1; GSPDB:GN001

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC0215

A:Map position: 1

C:Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 55.4%; Score 36; DB 2; Length 399;

Best Local Similarity 58.3%; Pred. No. 39;

Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 PLGVQRAQFQSG 13

Db 243 PVGVNRILVQSG 254

RESULT 8

S16582

fructose-bisphosphatase (EC 3.1.3.11) precursor, chloroplast - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 21-Nov-1993 #sequence_revision 12-May-1995 #text_change 22-Jun-1999

C:Accession: S16582

R:Horsnell, P.R.; Raines, C.A.

Plant Mol. Biol. 17, 185-186, 1991

A:Title: Nucleotide sequence of a cDNA clone encoding chloroplast fructose-1,6-bisphosph

A:Reference number: S16582; MUID:91329733; PMID:1651131

A:Accession: S16582

A:Molecule type: mRNA

A:Residues: 1-417 <HOR>

A:Cross-references: EMBL:X56148; NID:gl1241; PIDN:CAA41154.1; PID:gl1242

A:Experimental source: Clone AFBP1

C:Genetics:

A:Genome: nuclear

C:Superfamily: fructose-bisphosphatase

C:Keywords: Calvin cycle; chloroplast; gluconeogenesis; phosphoric monoester hydrolase

F;1-58/Domain: transit peptide (chloroplast) #status predicted <INP>

F;59-417/Product: fructose-bisphosphatase #status predicted <MA>

Query Match 55.4%; Score 36; DB 2; Length 417;

Best Local Similarity 53.8%; Pred. No. 41;

Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 APLGVQRAQFQSG 13

Db 43 APTGVNRHQYASG 55

RESULT 9

T47564

fructose-bisphosphatase precursor - Arabidopsis thaliana

N:Alternate names: protein F24B22.10

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 02-Sep-2000

C:Accession: T47564

R:Bioecker, H.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Queirer, F.; Salanoubat, M.

submitted to the Protein Sequence Database, January 2000

A:Reference number: Z23016

A:Accession: T47564

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-417 <BLO>

A:Cross-references: EMBL:AL132957

A:Experimental source: cultivar Columbia; BAC clone F24B22

C:Genetics:

A:Map position: 3

A:Introns: 149/3; 225/3; 394/3

A>Note: F24B22.10

C:Superfamily: fructose-bisphosphatase

Query Match 55.4%; Score 36; DB 2; Length 417;

Best Local Similarity 53.8%; Pred. No. 41;

Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 APLGVQRAQFQSG 13

Db 43 APTGVNRHQYASG 55

RESULT 10

T30938

receptor tyrosine phosphatase - medicinal leech

C:Species: Hirudo medicinalis (medicinal leech)

C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 31-Jan-2000

C:Accession: T30938

R:Gershon, T.R.; Baker, M.; Nitabach, M.; Wu, P.; Macagno, E.R.

submitted to the EMBL Data Library, December 1997

A:Description: Two receptor tyrosine phosphatases of the LAR subfamily are expressed

A:Reference number: Z20939

A:Accession: T30938

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2051 <GER>

A:Cross-references: EMBL:AF017083; NID:G2695654; PID:G2695655; PIDN:AAE91460.1

C:Genetics:

A:Gene: LAR2

C:Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homology

QY

Query Match 55.4%; Score 36; DB 2; Length 2051;

Best Local Similarity 63.6%; Pred. No. 2.1e+02;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 PLGVQRAQFQSG 12

Db 1439 FVDVRRMQFQSG 1449

RESULT 11

AC0990

conserved hypothetical protein STY4228 [imported] - Salmonella enterica subsp. enteri

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A>Note: this species has also been called Salmonella typhi

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: AC0990

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church

th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr

, S.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001.
 A;Authors: Parry, C.; Quall, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
 A;Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov
 A;Reference number: AB0502; MUID:21534947; PMID:11677608
 A;Accession: AC0990
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-192 <PAR>
 A;Cross-references: GB:AL513382; PIDN:CAD08047.1; PID:gl6505027; GSPDB:GN00176
 C;Genetics:
 A;Gene: STY4228

Query Match 53.8%; Score 35; DB 2; Length 192;
 Best Local Similarity 53.8%; Pred. No. 28;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 APLGVORAQFQSG 13
 ||||| :|||
 Db 25 APOGVRASWLAG 37

RESULT 12
 G75476
 conserved hypothetical protein - *Deinococcus radiodurans* (strain R1)
 C;Species: *Deinococcus radiodurans*
 C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000

A;Accession: G75476
 R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
 M.; Shen, M.; Vanathavan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999

A;Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.
 A;Reference number: AV5250; MUID:20036836; PMID:10567266

A;Accession: G75476
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-237 <WHI>
 A;Cross-references: GB:AE001933; GB:AE000513; NID:G6458490; PIDN:AAF10364.1; PID:G645849
 A;Experimental source: strain R1
 C;Genetics:
 A;Gene: DR0787
 A;Map position: 1

Query Match 53.8%; Score 35; DB 2; Length 237;
 Best Local Similarity 70.0%; Pred. No. 35;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 GVQRAQFQSG 13
 ||||| :|||
 Db 173 GVSRAVFQOG 182

RESULT 13
 T31699
 hypothetical protein F29A7.4 - *Caenorhabditis elegans*
 C;Species: *Caenorhabditis elegans*
 C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 R;Waterston, R.; Le, T.T.
 submitted to the EMBL Data Library, July 1997
 A;Description: The sequence of *C. elegans* cosmid F29A7.
 A;Reference number: Z21071
 A;Accession: T31699
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: DNA
 A;Residues: 1-268 <WAT>
 A;Cross-references: EMBL:AF016416; PIDN:AAB65274.1; GSPDB:GN00020; CESP:F29A7.4
 A;Experimental source: strain Bristol N2; clone F29A7
 C;Genetics:
 A;Gene: CESP:F29A7.4
 A;Map position: 2
 A;Ictrons: 21/1; 76/2; 121/2; 183/3; 202/3; 211/2; 234/3

Query Match 53.8%; Score 35; DB 2; Length 268;
 Best Local Similarity 58.3%; Pred. No. 40;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 APLGVORAQFQSG 12
 ||||| :|||
 Db 167 APEGOMKAEFQS 178

RESULT 14
 S43876
 site-specific DNA methyltransferase (EC 2.1.1.-) - *Caulobacter crescentus*
 C;Species: *Caulobacter crescentus*
 C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 A;Accession: S43876
 R;Zweiger, G.; Marczyński, G.; Shapiro, L.
 J. Mol. Biol. 235, 472-485, 1994
 A;Title: A *Caulobacter* DNA methyltransferase that functions only in the predivisional
 A;Reference number: S43876; MUID:94118303; PMID:8289276
 A;Accession: S43876
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-358 <ZWE>
 A;Cross-references: EMBL:U01032; NID:G393011; PIDN:AAI8913.1; PID:G393012
 C;Superfamily: site-specific DNA-methyltransferase (adenine-specific) HpaI
 C;Keywords: methyltransferase; restriction modification system

Query Match 53.8%; Score 35; DB 1; Length 358;
 Best Local Similarity 53.8%; Pred. No. 54;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 APLGVORAQFQSG 13
 ||||| :|||
 Db 344 APIDVLRQVRAG 356

RESULT 15
 A87296
 modification methylase CcrMI [imported] - *Caulobacter crescentus*
 C;Species: *Caulobacter crescentus*
 C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 17-May-2002
 C;Accession: A87296
 R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gilm, M.L.; Haft, D.H.; K
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser,
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A;Title: Complete Genome Sequence of *Caulobacter crescentus*.
 A;Reference number: A87249; MUID:21173698; PMID:11259647
 A;Accession: A87296
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-358 <STO>
 A;Cross-references: GB:AE005673; NID:gl3421535; PIDN:AAK22365.1; GSPDB:GN00148
 C;Genetics:
 C;Superfamily: site-specific DNA-methyltransferase (adenine-specific) HpaI

Query Match 53.8%; Score 35; DB 2; Length 358;
 Best Local Similarity 53.8%; Pred. No. 54;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 APLGVORAQFQSG 13
 ||||| :|||
 Db 344 APIDVLRQVRAG 356

Query Match 53.8%; Score 35; DB 2; Length 358;
 Best Local Similarity 53.8%; Pred. No. 54;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 APLGVORAQFQSG 13
 ||||| :|||
 Db 344 APIDVLRQVRAG 356

Search completed: January 20, 2004, 17:03:38
 Job time : 15.4962 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:41:41 ; Search time 6.84733 Seconds
(without alignments)
89.283 Million cell updates/sec

Title: US-10-089-549-18
Perfect score: 65

Sequence: 1 APLGVQRAQFQSG 13

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	60.0	2245	MYSJ_DICDI	P54697 dictyosteli
2	39	60.0	3898	POLG_BVDVS	Q01499 bovine vira
3	39	60.0	3988	POLG_BVDVN	P19711 bovine vira
4	36	55.4	417	F16P_ARATH	P25851 arabidopsis
5	25	53.8	132	ACPT_SALT1	Q82259 salmonella
6	35	53.8	132	ACPT_SALT2	Q82259 salmonella
7	35	53.8	358	MTC1_CAUCR	Q45971 caulobacter
8	35	53.8	660	VST2_HEVBU	P29326 hepatitis e
9	35	53.8	660	VST2_HEVPA	Q04611 hepatitis e
10	35	53.8	732	DD35_CAEEL	P33426 hepatitis e
11	35	53.8	732	SYFB_THRTH	Q9bkq8 caenorhabdi
12	35	53.8	785	YAT2_SYNP6	P27002 thermus the
13	34	52.3	208	SP4_EACME	P41028 bacillus me
14	34	52.3	247	YBGF_ECO57	Q8x6q5 escherichia
15	34	52.3	375	YBGF_ECOLI	P75817 escherichia
16	34	52.3	375	PHP5_SCHPO	P79007 schizosacch
17	34	52.3	415	PKN1_MYXXA	P33973 myxococcus
18	34	52.3	693	BGL1_SACFI	P22506 saccharomyc
19	34	52.3	876	UREB_KLEPN	Q02944 klebsiella
20	33	50.8	270	DJLA_ECOLI	P31680 escherichia
21	33	50.8	271	YCUQ_ECOLI	P76043 escherichia
22	33	50.8	350	WNT6_MOUSE	P22727 mus musculu
23	33	50.8	364	WNT6_HUMAN	Q9y6f9 homo sapien
24	33	50.8	365	APF1_HUMAN	P53367 homo sapien
25	33	50.8	373	GCH2_ARCFU	Q29766 archaeglob
26	33	50.8	378	YV80_MYCTU	O06343 mycobacteri
27	33	50.8	497	FXH1_XENLA	P70056 xenopus lae
28	33	50.8	518	IPGD_SHIFL	Q07566 shigella fl
29	33	50.8	538	IPGD_SHISO	Q55286 shigella so
30	33	50.8	538	PTMA_PASMU	Q9c1y8 pasteurella
31	33	50.8	624	BDF1_YEAST	P35817 saccharomyc
32	33	50.8	686	BDF1_YEAST	P35817 saccharomyc
33	33	50.8	880	BGL2_SACFI	P22507 saccharomyc

34 33 50.8 1048 1 SBCC_ECOLI
35 33 50.8 1071 1 PR16_YEAST
36 33 50.8 1127 1 Y855_TREPA
37 33 50.8 1173 1 GIGA_ORISA
38 33 50.8 5065 1 EPPL_HUMAN
39 32 49.2 116 1 RL20_MYCPU
40 32 49.2 153 1 FABZ_CHLUPN
41 32 49.2 153 1 FABZ_CHLUPN
42 32 49.2 153 1 FABZ_CHLTR
43 32 49.2 156 1 GREY_THEMA
44 32 49.2 208 1 KTHY_CAUCR
45 32 49.2 210 1 NADD_STRP3

ALIGNMENTS

RESULT 1

MYSJ_DICDI
ID MYSJ_DICDI STANDARD; PRT; 2245 AA.
AC P54697;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Myosin IJ heavy chain.
DE MFOJ.
GN Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RX MEDLINE=96215148; PubMed=8636147;
RA Hammer J.A. III, Jung G.;
RT "The sequence of the dictyostelium myo J heavy chain gene predicts a novel, dimeric, unconventional myosin with a heavy chain molecular mass of 258 kDa."
RT J. Biol. Chem. 271:7120-7127(1996).
RN [2]
RP SEQUENCE OF 1-1021 FROM N.A.
RX MEDLINE=97039016; PubMed=884597;
RA Peterson M.D., Urioste A.S., Titus M.A.;
RT "Dictyostelium discoideum myoJ: a member of a broadly defined myosin V class or a class XI unconventional myosin?";
RL J. Muscle Res. Cell Motil. 17:411-424(1996).
RN [3]
RP SEQUENCE OF 182-298 FROM N.A.
RX MEDLINE=95023928; PubMed=7937787;
RA Titus M.A., Kuspa A., Loomis W.F.;
RT "Discovery of myosin genes by physical mapping in Dictyostelium.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:9446-9450(1994).
CC -!- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN AND HAS ATPASE ACTIVITY THAT IS ACTIVATED BY ACTIN.
CC -!- SUBUNIT: Homodimer.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 dilute domain.
CC -!- SIMILARITY: Contains 3 IQ domains.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U42409; AAA85186.1; --
EMBL; L35322; AAA79858.1; --
PIR; T18278; T18278.
DR HSSP; P08799; 1MND.
DR DictyDB; D001095; myoJ.
DR InterPro; IPR002710; DIL.
DR InterPro; IPR000048; IQ_region.

```

DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR004009; Myosin_N.
DR Pfam; PF01843; DIL; 1.
DR Pfam; PF00612; IQ; 6.
DR Pfam; PF00063; myosin_head; 2.
DR Pfam; PF02736; Myosin_N.1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD003376; DIL; 1.
DR SMART; SM00015; IQ; 3.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS00096; IQ; 3.
DR Myosin; ATP-binding; Actin-binding; Calmodulin-binding; Repeat;
KW Coiled coil.
FT DOMAIN 1 809 MYOSIN HEAD-LIKE.
FT DOMAIN 824 851 IQ 1.
FT DOMAIN 872 901 IQ 2.
FT DOMAIN 943 972 IQ 3.
FT DOMAIN 973 1812 COILED COIL (POTENTIAL).
FT DOMAIN 2560 2220 DILUTE.
FT NP_BIND 174 181 ATP (POTENTIAL).
FT DOMAIN 669 749 MYOSIN HEAD-LIKE.
FT CONFLICT 191 191 L -> F (IN REF. 2).
FT CONFLICT 284 284 A -> T (IN REF. 2).
FT CONFLICT 291 291 G -> R (IN REF. 2).
FT CONFLICT 332 347 NKGCFEIEGVSDDEH -> IEWMFELKVRWKS (IN REF. 2).
FT CONFLICT 550 550 N -> K (IN REF. 2).
FT CONFLICT 865 866 HH -> QQ (IN REF. 2).
FT SEQUENCE 2245 AA; 258478 MW; 61585EFDIAB45BE CRC64;

Query Match 60.0%; Score 39; DB 1; Length 2245;
Best Local Similarity 72.7%; Pred. No. 23;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 3 LGVRAQFQSG 13
Db 1190 LGQSSQFQSG 1200

RESULT 2
ID POLG BVDVNS STANDARD; PRT; 3898 AA.
AC Q01499;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Genome polyprotein.
OS Bovine viral diarrhoea virus (strain SD-1) (BVDV) (Mucosal disease virus).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Pestivirus.
OX NCBI_TaxID=31656;
RN [1]
RP MEDLINE=93079889; PubMed=1333126;
RA Deng R.; Brock K.V.;
RT nontyphoidic bovine viral diarrhoea virus strain SD-1.;
RL Virology 191:867-869(1992).
CC -!- FUNCTION: PESTIVIRUS P80 (P125) MAY BE A BIFUNCTIONAL PROTEIN WITH HELICASE AND PROTEASE ACTIVITY.
CC -!- PTM: GP116 GIVES RISE TO GP62 AND GP53; GP62 IN TURN YIELDS GP48 AND GP25.
CC -!- SIMILARITY: TO THE HOG CHOLERA VIRUS GENOME POLYPROTEIN.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S31.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```

```

CC or send an email to license@isb-sib.ch.
CC
DR EMBL; M96751; AAA2860.1; -.
DR PIR; A44217; A44217.
DR HSP; P27958; IALV.
DR MEROPS; C33.001; -.
DR MEROPS; S31.001; -.
DR InterPro; IPR000280; CDvir_endptsp80.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002166; HCV RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007895; RNA_pol_Bs_Ps.
DR InterPro; IPR007094; RNA_pol_Pevir.
DR InterPro; IPR001568; RNase_T2.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral RdRP; 1.
DR PRINTS; PR00729; CDVENDOPTASE.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICCC; 1.
DR PROSITE; PS00531; RNASE_T2_2; UNKNOWN_1.
DR Polyprotein; Glycoprotein; Helicase; Serine protease; Hydrolase.
FT CHAIN 1 2270 P20 (30KD).
FT CHAIN 2271 21063 GP116/GP62-GP53 (GLYCOPROTEIN).
FT CHAIN ? ? GP125/GP54-GP80.
FT CHAIN ? ? GP133/GP58-GP75.
FT DOMAIN 690 755 CYS-RICH.
FT ACT_SITE 1658 1658 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1695 1695 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1752 1752 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 272 272 N-LINKED (GLCNAC).
FT CARBOHYD 281 281 N-LINKED (GLCNAC).
FT CARBOHYD 296 296 N-LINKED (GLCNAC).
FT CARBOHYD 335 335 N-LINKED (GLCNAC).
FT CARBOHYD 365 365 N-LINKED (GLCNAC).
FT CARBOHYD 370 370 N-LINKED (GLCNAC).
FT CARBOHYD 413 413 N-LINKED (GLCNAC).
FT CARBOHYD 487 487 N-LINKED (GLCNAC).
FT CARBOHYD 597 597 N-LINKED (GLCNAC).
FT CARBOHYD 809 809 N-LINKED (GLCNAC).
FT CARBOHYD 878 878 N-LINKED (GLCNAC).
FT CARBOHYD 922 922 N-LINKED (GLCNAC).
FT CARBOHYD 990 990 N-LINKED (GLCNAC).
FT CARBOHYD 1357 1357 N-LINKED (GLCNAC).
FT CARBOHYD 1419 1419 N-LINKED (GLCNAC).
FT CARBOHYD 1451 1451 N-LINKED (GLCNAC).
FT CARBOHYD 1713 1713 N-LINKED (GLCNAC).
FT CARBOHYD 2134 2134 N-LINKED (GLCNAC).
FT CARBOHYD 2217 2217 N-LINKED (GLCNAC).
FT CARBOHYD 2494 2494 N-LINKED (GLCNAC).
FT CARBOHYD 2682 2682 N-LINKED (GLCNAC).
FT CARBOHYD 2751 2751 N-LINKED (GLCNAC).
FT CARBOHYD 2891 2891 N-LINKED (GLCNAC).
FT CARBOHYD 2988 2988 N-LINKED (GLCNAC).
FT CARBOHYD 3688 3688 N-LINKED (GLCNAC).
FT CARBOHYD 3777 3777 N-LINKED (GLCNAC).
FT CARBOHYD 3793 3793 N-LINKED (GLCNAC).
FT SEQUENCE 3898 AA; 437800 MW; 31ACBE140D407ED3 CRC64;

Query Match 60.0%; Score 39; DB 1; Length 3898;
Best Local Similarity 70.0%; Pred. No. 42;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 4 GVGRAQFQSG 13
Db 285 GIGRAMFORG 294

RESULT 3
POLG BVDVNS STANDARD; PRT; 3988 AA.
ID POLG BVDVNS STANDARD; PRT; 3988 AA.
AC P19711;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)

```

FT	CARBOHYD	1357	1357	N-LINKED (GLUCNA	. .)	(POTENTIAL)
FT	CARBOHYD	1419	1419	N-LINKED (GLUCNA	. .)	(POTENTIAL)
FT	CARBOHYD	1451	1451	N-LINKED (GLUCNA	. .)	(POTENTIAL)
FT	CARBOHYD	1803	1803	N-LINKED (GLUCNA	. .)	(POTENTIAL)
FT	CARBOHYD	2224	2224	N-LINKED (GLUCNA	. .)	(POTENTIAL)
FT	CARBOHYD	2307	2307	N-LINKED (GLUCNA	. .)	(POTENTIAL)
FT	CARBOHYD	2584	2584	N-LINKED (GLUCNA	. .)	(POTENTIAL)
FT	CARBOHYD	2772	2772	N-LINKED (GLUCNA	. .)	(POTENTIAL)
FT	CARBOHYD	2981	2981	N-LINKED (GLUCNA	. .)	(POTENTIAL)
FT	CARBOHYD	3778	3778	N-LINKED (GLUCNA	. .)	(POTENTIAL)
FT	CARBOHYD	3867	3867	N-LINKED (GLUCNA	. .)	(POTENTIAL)
FT	CARBOHYD	3883	3883	N-LINKED (GLUCNA	. .)	(POTENTIAL)
SQ	SEQUENCE	3988 AA;	449154 MW;	4474212F338661B8	CRC64;	

Query Match 60.0%; Score 39; DB 1; Length 3988;
Best Local Similarity 70.0%; Pred. No. 43;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps

Qy	4	GVORAQFQSG 13	
		: : : :	
Db	285	GIQRAMFQFG 294	

RESULT 4

F16P	ARATH	STANDARD;	PRT;	417 AA.
ID	F16P ARATH	Q9M398;		
AC	P25851;			
DT	01-MAY-1992	(Rel. 22, Created)		
DT	28-FEB-2003	(Rel. 41, Last sequence update)		
DT	28-FEB-2003	(Rel. 41, Last annotation update)		
DE	(D-fructose-1,6-bisphosphatase, chloroplast precursor (EC 3.1.3.11))			
DE	(D-fructose-1,6-bisphosphate 1-phosphohydrolase) (FBbase).			
GN	F5P OR AT3G54050 OR F24822.10.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosids II; Brassicales; Brassicaceae; Arabidopsis.			
OX	NCBI_TaxID=3702;			
ON	[1]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=91329733; PubMed=1651131;			
EX	Horsnell P.N., Raines C.A.;			
RT	"Nucleotide sequence of a cDNA clone encoding chloroplast			
RT	fructose-1,6-bisphosphatase from Arabidopsis thaliana.";			
RL	Plant Mol. Biol. 17:185-186(1991).			
RP	SEQUENCE FROM N.A.			
RP	STRALN=CV. Columbia;			
RX	MEDLINE=21016720; PubMed=11130713;			
RA	Salanoubat M., Lemcke K., Rieger M., Ansoerg W., Unseld M.,			
RA	Fartmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,			
RA	Delsen Y.M., Boutry M., Grievell L.A., Mache R., Puigdomenech P.,			
RA	De Simone V., Choise N., Artiguenave F., Robert C., Brottier F.,			
RA	Wincker P., Cattolico L., Weissensbach J., Saurin W., Quetier F.,			
RA	Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,			
RA	Wurmback E., Drzonek H., Erfle H., Jordan N., Bangert S.,			
RA	Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,			
RA	Vezi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,			
RA	Conrad A., Hornischer K., Kauer G., Loehert T.-H., Nordstiek G.,			
RA	Reichelt J., Scharfe M., Schoen O., Bargues M., Terol J., Clement J.,			
RA	Navearro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,			
RA	Cooke R., Laudie M., Berger-Lilauro C., Purnelle B., Maey D.,			
RA	de Haan M., Maere A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,			
RA	Monfort A., Argirou A., Flores M., Liguori R., Vitale D.,			
RA	Manhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,			
RA	Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,			
RA	Reichelt J., Scharfe M., Schoen O., Bargues M., Terol J., Clement J.,			
RA	Navearro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,			
RA	Cooke R., Laudie M., Berger-Lilauro C., Purnelle B., Maey D.,			
RA	de Haan M., Maere A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,			
RA	Monfort A., Argirou A., Flores M., Liguori R., Vitale D.,			
RA	Manhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,			
RA	Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,			
RA	Reichelt J., Scharfe M., Schoen O., Bargues M., Terol J., Clement J.,			
RA	Navearro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,			
RA	Cooke R., Laudie M., Berger-Lilauro C., Purnelle B., Maey D.,			
RA	de Haan M., Maere A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,			
RA	Monfort A., Argirou A., Flores M., Liguori R., Vitale D.,			
RA	Manhaupt G., Haase D., Schoof H.,			

RA Nakayama S., Nakazaki N., Shimpō S., Takeuchi C., Wada T.,
 RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
 RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
 RT italiana";
 RL Nature 408:820-822(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
 RT "RIKEN Arabidopsis full length cDNA clones (RAPFs) sequenced by the
 RT SSP consortium (Salk/Stanford/PECC)";
 RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: D-fructose 1,6-bisphosphate + H₂O = D-
 CC fructose 6-phosphate + phosphate.
 CC -!- PATHWAY: THE CHLOROPLAST ISOZYME TAKES PART IN THE REGENERATION OF
 CC RIBULOSE BIPHOSPHATE IN THE PHOTOSYNTHETIC CARBON REDUCTION
 CC CYCLE (CALVIN CYCLE).
 CC -!- SUBUNIT: Homotetramer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Chloroplast.
 CC -!- INDUCTION: LIGHT ACTIVATION THROUGH PH CHANGES, MG(2+) LEVELS
 CC AND ALSO BY LIGHT-MODULATED REDUCTION OF ESSENTIAL DISULFIDE
 CC GROUPS VIA THE FERREDOXIN-THIOREDOXIN F SYSTEM (BY SIMILARITY).
 CC -!- MISCELLANEOUS: IN PLANTS THERE ARE TWO FBPASE ISOZYMES: ONE IN THE
 CC CYTOSOL AND THE OTHER IN THE CHLOROPLAST.
 CC -!- SIMILARITY: BELONGS TO THE FBPASE FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL; X58148; CAA41154.1; -;
 DR EMBL; AL132957; CAB70973.1; -;
 DR EMBL; AF428326; AAL6256.1; -;
 DR EMBL; AY039934; AAK64038.1; -;
 DR PIR; S16582; S16582.
 DR PIR; T47564; T47564.
 DR HSP; P46275; IDBZ.
 DR InterPro; IPR000146; In_FB_phptase.
 DR Pfam; PF00316; FBPAse; 1.
 DR PRINTS; PR00377; INFBPHPTASE.
 DR ProDom; PD001491; In_FB_phptase; 1.
 DR PROSITE; PS00124; FBPAse; 1.
 KW Hydrolyase; Carbohydrate metabolism; Multigene family; Chloroplast;
 KW Transit peptide; Calvin cycle.
 FT TRANSIT 1 59
 FT CHAIN 60 417
 FT ACT SITE 359 359
 FT DISULFID 233 238
 FT CONFLICT 4 4
 FT CONFLICT 66 66
 FT CONFLICT 133 133
 FT CONFLICT 133 133
 SQ SEQUENCE 417 AA; 45162 MW; CEB17F65468746D7 CRC64;
 Query Match 55.4%; Score 36; DB 1; Length 417;
 Best Local Similarity 53.8%; Pred. No. 16;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 3;
 QY 1 APLGVQRAQFQSG 13
 Db 43 APTGVNRHQYASG 55
 RESULT 5
 ACPT SALT
 ID ACPT SALT
 AC Q82259;
 DT 28-FEB-2003 (Rel. 41, Created)
 DE 4'-phosphopantetheinyl transferase acpt (EC 2.7.8.-).
 GN ACPT OR STM3583.
 OS Salmonella typhimurium.

DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE 4'-phosphopantetheinyl transferase acpt (EC 2.7.8.-).
 GN ACPT OR STV4228 OR T3939.
 OS Salmonella typhi.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Salmonella.
 CC NCBI TaxID=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Dickard D., Wain J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Conerton P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Felwell T., Hamlin N., Hogue A., Hien T.T., Holroyd S., Jagels K.,
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrell B.G.;
 RT "Complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhi CT18";
 RL Nature 413:848-852(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Ty2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
 RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
 RT and CT18";
 RL J. Bacteriol. 185:2330-2337(2003).
 CC -!- FUNCTION: May be involved in an alternative pathway for
 CC phosphopantetheinyl transfer and holo-ACP synthesis. The
 CC native apo-protein substrate is unknown (By similarity).
 CC -!- CATALYTIC ACTIVITY: CoA + [protein-X] = adenosine 3',5'-
 CC bisphosphate + phosphopantetheinyl-[protein-X].
 CC -!- SIMILARITY: BELONGS TO THE P-PANT TRANSFERASE SUPERFAMILY.
 CC GSP/SFP/HETI/ACPT FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL; AL627281; CAD08047.1; -;
 DR EMBL; AE016847; AA071410.1; -;
 DR InterPro; IPR004568; Pantethn_trn.
 DR Pfam; PF01648; ACPS; 1.
 KW Transferase; Complete proteome.
 SQ SEQUENCE 192 AA; 20871 MW; C1C257B1707141F3 CRC64;
 Query Match 53.8%; Score 35; DB 1; Length 192;
 Best Local Similarity 53.8%; Pred. No. 11;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 1 APLGVQRAQFQSG 13
 Db 25 APQGVRRASWLAG 37
 RESULT 6
 ACPT SALT
 ID ACPT SALT
 AC Q8ZLE2;
 DT 28-FEB-2003 (Rel. 41, Created)
 DE 4'-phosphopantetheinyl transferase acpt (EC 2.7.8.-).
 GN ACPT OR STM3583.
 OS Salmonella typhimurium.

```

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan X.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856 (2001).
CC -!- FUNCTION: May be involved in an alternative pathway for
CC phosphopantetheinyl transfer and holo-ACP synthesis. The
CC native apo-protein substrate is unknown (By similarity).
CC -!- CATALYTIC ACTIVITY: CoA + [protein-X] = adenosine 3',5'-
CC bisphosphate + phosphopantetheinyl-[protein-X].
CC -!- SIMILARITY: BELONGS TO THE P-PANT TRANSFERASE SUPERFAMILY.
CC GSP/SFP/HETI/ACPT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE008845; AAL22443.1; -.
DR StyGene; SG????; acpt.
DR InterPro; IPR004568; Pantethn_trn.
DR Pfam; PF01648; ACPs; 1.
KW Transferase; Complete proteome.
SQ SEQUENCE 192 AA; 20799 MW; E530B7A3BF30EC7A CRC64;

Query Match 53.8%; Score 35; DB 1; Length 192;
Best Local Similarity 53.8%; Pred. No. 11;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 APLGVQRAQFQSG 13
|||:||||:|
Db 25 APQGVRRASWLAG 37

RESULT 7
ID _MTC1_CAUCR STANDARD; PRT; 358 AA.
AC Q4597L;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Modification methylase Ccrml [EC 2.1.1.72] (Adenine-specific
DE methyltransferase Ccrml) (M.Ccrml).
GN CCRMIM OR CCRM OR CC0378.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=CB15N / NA1000;
RX MEDLINE=94118303; PubMed=8289276;
RA Zweiger G., Marczynski G., Shapiro L.;
RT "A Caulobacter DNA methyltransferase that functions only in the
RT predivisional cell."
RL J. Mol. Biol. 235:472-485 (1994).
RN [2]
RC SEQUENCE FROM N.A.
RX STRAIN=ATCC 19069 / CB15;
RX MEDLINE=21173698; PubMed=11259647;

```

```

RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Shadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ertolaeva M., White O.,
RA Salzberg S.B., Venter J.C., Shapiro L., Fraser C.W.;
RT "Complete genome sequence of Caulobacter crescentus."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141 (2001).
CC -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE
CC GANTC AND CAUSES SPECIFIC METHYLATION ON A-2 ON BOTH STRANDS.
CC CCRM-MEDIATED METHYLATION HAS IMPORTANT CELLULAR FUNCTIONS.
CC APPEARS TO CONTRIBUTE TO THE ACCURATE CELL-CYCLE CONTROL OF DNA
CC REPLICATION AND CELLULAR MORPHOLOGY.
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
CC adenosyl-L-homocysteine + DNA 6-methylaminopurine.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U01032; AAA18913.1; -.
DR EMBL; AE005711; AAK22365.1; -.
DR PIR; A87296; A87296.
DR PIR; S43876; S43876.
DR HSP; P11409; 1BOO.
DR REBASE; 2539; M.Ccrml.
DR TIGR; CC0378; -.
DR InterPro; IPR001091; CNA_Mettransf.
DR InterPro; IPR002295; D21N6_mtfase.
DR PIR; A87296; A87296.
DR InterPro; IPR002941; N6/N4_Mtase.
DR InterPro; IPR002052; N6_Mtase.
DR InterPro; IPR000051; SAM_bind.
DR Pfam; PF01555; N6_N4_Mtase; 1.
DR PRINTS; PR00506; D21N6MTFRASE.
DR PRINTS; PR00508; S21N4MTFRASE.
DR PROSITE; PS00092; N6_MTASE; 1.
KW Transferase; Methyltransferase; DNA replication; Complete proteome.
FT CONFLICT 242 242 05F43266F7D4C614 CRC64;
SQ SEQUENCE 358 AA; 39665 MW; 05F43266F7D4C614 CRC64;

Query Match 53.8%; Score 35; DB 1; Length 358;
Best Local Similarity 53.8%; Pred. No. 22;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 APLGVQRAQFQSG 13
|||:||||:|
Db 344 APIDVIRAQVRAG 356

RESULT 8
ID VST2_HEVBU STANDARD; PRT; 660 AA.
AC P29326;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Structural protein 2 precursor (ORF2).
OS Hepatitis E virus (strain Burma) (HEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage;
OC Hepatitis E-like viruses.
OX NCBI_TaxID=31767;
RN [1]
RC SEQUENCE FROM N.A.
RX MEDLINE=92024067; PubMed=1926770;
RA Tam A.W., Smith M.M., Guerra M.E., Huang C.-C., Bradley D.W.,
RA Fry K.E., Reyes G.R.;
RT "Hepatitis E virus (HEV): molecular cloning and sequencing of the
RT full-length viral genome."

```

RL Virology 185:120-131(1991).
 CC -!- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING
 CC THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA
 CC BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M73218; AAA45736.1; -
 DR PIR; C40778; VHWWH2.
 DR InterPro; IPR004261; SP2.
 DR Pfam; PF03014; SP2; 1.
 KW Signal.
 FT SIGNAL. 1 19 POTENTIAL.
 FT CHAIN 20 660 STRUCTURAL PROTEIN 2.
 SQ SEQUENCE 660 AA; 70978 MW; 5832A013CCC4A61C CRC64;
 Query Match 53.8%; Score 35; DB 1; Length 660;
 Best Local Similarity 63.6%; Pred. No. 41;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 2 PLGVQRAQFQS 12
 Db 632 PLGLQCAQFS 642

 RESULT 9
 VST2 HEVMY
 ID VST2 HEVMY STANDARD; PRT; 660 AA.
 AC Q04611;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Structural protein 2 precursor (ORF2).
 OS Hepatitis E virus (strain Myanmar) (HEV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage;
 OC Hepatitis E-like viruses.
 OC NCBI_TaxID=31769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93227573; PubMed=8470371;
 RA Aye T.T., Uchida T., Ma M.Z., Lida F., Shikata T., Ichikawa M.,
 RA Rikihisa T., Winn K.;
 RT "Sequence and gene structure of the hepatitis E virus isolated from
 RT Myanmar.";
 RL Virus Genes 7:95-109(1993).
 CC -!- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING
 CC THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA
 CC BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D10330; BAA01174.1; -
 DR InterPro; IPR004261; SP2.
 DR Pfam; PF03014; SP2; 1.
 KW Signal.
 FT SIGNAL. 1 22 BY SIMILARITY.
 FT CHAIN 23 660 STRUCTURAL PROTEIN 2.
 SQ SEQUENCE 660 AA; 70998 MW; 3A82A4EA255C6253 CRC64;
 Query Match 53.8%; Score 35; DB 1; Length 660;
 Best Local Similarity 63.6%; Pred. No. 41;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 2 PLGVQRAQFQS 12
 Db 632 PLGLQCAQFS 642

 RESULT 10
 VST2 HEVPA
 ID VST2 HEVPA STANDARD; PRT; 660 AA.
 AC P33426;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE Structural protein 2 precursor (ORF2).
 OS Hepatitis E virus (strain Pakistan) (HEV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage;
 OC Hepatitis E-like viruses.
 OC NCBI_TaxID=33774;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92115700; PubMed=1731327;
 RA Tearev S.A., Emerson S.U., Reyes G.R., Teareva T.S., Legters L.J.,
 RA Malik I.A., Iqbal M., Purcell R.H.;
 RT "Characterization of a prototype strain of hepatitis E virus.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:559-563(1992).
 CC -!- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING
 CC THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA
 CC BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M80581; AAA45727.1; -
 DR InterPro; IPR004261; SP2.
 DR Pfam; PF03014; SP2; 1.
 KW Signal.
 FT SIGNAL. 1 22 BY SIMILARITY.
 FT CHAIN 23 660 STRUCTURAL PROTEIN 2.
 SQ SEQUENCE 660 AA; 70980 MW; 8085BC53CFB46FD3 CRC64;
 Query Match 53.8%; Score 35; DB 1; Length 660;
 Best Local Similarity 63.6%; Pred. No. 41;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 2 PLGVQRAQFQS 12
 Db 632 PLGLQCAQFS 642

 RESULT 11
 DD35 CAEEL
 ID DD35 CAEEL STANDARD; PRT; 732 AA.
 AC Q9BKQ8;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable ATP-dependent helicase DBX35 homolog.
 GN Y67D2.6.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OC NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Waterston R.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

```

CC CC  -!- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY. DEAH
CC CC  SUBFAMILY.
CC CC  -----
CC CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC  the European Bioinformatics Institute. There are no restrictions on its
CC CC  use by non-profit institutions as long as its content is in no way
CC CC  modified and this statement is not removed. Usage by and for commercial
CC CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC  or send an email to license@isb-sib.ch).
CC CC  -----
CC CC  EMBL: AC091125; AKX27888.1; -.
CC CC  Wormpep; Y67D2.6; CE27311.
CC CC  InterPro: IPR001410; DEAD.
CC CC  InterPro: IPR002464; DEAH_Box.
CC CC  InterPro: IPR001650; Helicase_C.
CC CC  Pfam: PF04408; HA2; 1.
CC CC  Pfam: PF00271; Helicase_C; 1.
CC CC  SMART: SM00487; DEXDC; 1.
CC CC  SMART: SM00490; HELIC; 1.
CC CC  DR PROSITE: PS00690; DEAH ATP HELICASE; 1.
CC CC  KW Hypothetical protein; Hydrolase; Helicase; ATP-binding.
CC CC  FT NP_BIND 100 107 ATP (POTENTIAL).
CC CC  FT SITE 198 201 DEAH BOX.
CC CC  SQ SEQUENCE 732 AA; 82543 MW; E64171070E43BF41 CRC64;
CC CC  Query Match 53.8%; Score 35; DB 1; Length 732;
CC CC  Best Local Similarity 60.0%; Pred. No. 46;
CC CC  Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
CC CC  -----
CC CC  1 APLGVORACQF 10
CC CC  :|||:|:|:
CC CC  496 SPLGLQWAEF 505
CC CC  -----
CC CC  RESULT 12
CC CC  SYFB_THETH STANDARD; PRT; 785 AA.
CC CC  ID SYFB_THETH STANDARD; PRT; 785 AA.
CC CC  AC P27002;
CC CC  DT 01-AUG-1992 (Rel. 23, Created)
CC CC  DT 01-DEC-1992 (Rel. 24, Last sequence update)
CC CC  DT 15-SEP-2003 (Rel. 42, Last annotation update)
CC CC  DE Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20)
CC CC  DE (Phenylalanine--tRNA ligase beta chain) (PheRS).
CC CC  GN PHET.
CC CC  OS Thermus thermophilus.
CC CC  OS Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
CC CC  OS Thermus.
CC CC  OX NCBI_TaxID=274;
CC CC  RN [1]
CC CC  RP SEQUENCE FROM N.A.
CC CC  RC STRAIN-HB8 / ATCC 27634;
CC CC  RX MEDLINE=93083630; PubMed=14511792;
CC CC  RA Keller B., Kast P., Hennecke H.;
CC CC  RT "Cloning and sequence analysis of the phenylalanyl-tRNA synthetase
CC CC  RT genes (phest) from Thermus thermophilus.";
CC CC  RL FEBS Lett. 301:83-88(1992).
CC CC  RN [2]
CC CC  RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
CC CC  RC STRAIN-HB8 / ATCC 27634;
CC CC  RX MEDLINE=92137522; PubMed=1508711;
CC CC  RA Kreutzer R., Kluft V., Bobkova E.V., Lavrik O.J., Sprinzl M.;
CC CC  RT "Structure of the phenylalanyl-tRNA synthetase genes from Thermus
CC CC  RT thermophilus HB8 and their expression in Escherichia coli.";
CC CC  RL Nucleic Acids Res. 20:4173-4178(1992).
CC CC  RN [3]
CC CC  RP SEQUENCE FROM N.A.
CC CC  RC STRAIN-HB8 / ATCC 27634;
CC CC  RX MEDLINE=98263257; PubMed=9600851;
CC CC  RA Lechner A., Kreutzer R.;
CC CC  RT "The phenylalanyl-tRNA synthetase specifically binds DNA.";
CC CC  RL J. Mol. Biol. 278:897-901(1998).
CC CC  RN [4]

```

```

RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS).
RC STRAIN-HB8 / ATCC 27634;
RX MEDLINE=94257735; PubMed=8199244;
RA Mosyak L., Sapiro M.;
RT "phenylalanyl-tRNA synthetase from Thermus thermophilus has four
RT antiparallel folds of which only two are catalytically functional.";
RL Biochimie 75:1091-1098(1993).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS).
RC STRAIN-HB8 / ATCC 27634;
RX MEDLINE=95393238; PubMed=7664121;
RA Mosyak L., Reshetnikova L., Goldgur Y., Delarue M., Sapiro M.G.;
RT "Structure of phenylalanyl-tRNA synthetase from Thermus
RT thermophilus.";
RL Nat. Struct. Biol. 2:537-547(1995).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RC MEDLINE=99194858; PubMed=10092459;
RA Reshetnikova L., Moor N., Lavrik O., Vassilyev D.G.;
RT "Crystal structures of phenylalanyl-tRNA synthetase complexed with
RT phenylalanine and a phenylalanyl-adenylate analogue.";
RL J. Mol. Biol. 287:555-568(1999).
CC -!- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +
CC diphosphate + L-phenylalanyl-tRNA(Phe).
CC -!- COFACTOR: Binds 2 magnesium ions per tetramer.
CC -!- SUBUNIT: Tetramer of two alpha and two beta chains.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the phenylalanyl-tRNA synthetase beta chain
CC family. Subfamily 1.
CC -!- SIMILARITY: Contains 1 tRNA-binding domain.
CC -----
CC CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC  the European Bioinformatics Institute. There are no restrictions on its
CC CC  use by non-profit institutions as long as its content is in no way
CC CC  modified and this statement is not removed. Usage by and for commercial
CC CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC  or send an email to license@isb-sib.ch).
CC CC  -----
CC CC  EMBL: Z121118; CAA78105.1; -.
CC CC  EMBL: X65609; CAA46560.1; -.
CC CC  EMBL: Y15464; CAA75645.1; -.
CC CC  PIR: T52503; T52503.
CC CC  PDB: 1PXS; 19-NOV-97.
CC CC  PDB: 1B70; 09-FEB-00.
CC CC  PDB: 1B7Y; 26-JAN-00.
CC CC  PDB: 1E1Y; 14-DEC-01.
CC CC  PDB: 1JUC; 02-NOV-01.
CC CC  HAMAP: MF_00283; -. 1.
CC CC  InterPro: IPR005146; B3_4.
CC CC  InterPro: IPR005147; B5.
CC CC  InterPro: IPR005121; Fdx-AntiCB.
CC CC  InterPro: IPR004532; PheT_bact.
CC CC  InterPro: IPR002319; tRNA-synt_2d.
CC CC  InterPro: IPR002547; tRNA_bind.
CC CC  Pfam: PF03483; B3_4; 1.
CC CC  Pfam: PF03484; B5; 1.
CC CC  Pfam: PF03147; FDX-ACB; 1.
CC CC  Pfam: PF01588; tRNA_bind; 1.
CC CC  TIGRfams: TIGR00472; pheT_bact; 1.
CC CC  PROSITE: PS00886; TRBD; 1.
CC CC  KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC CC  KW Metal-binding; Magnesium; RNA-binding; tRNA-binding; 3D-structure.
CC CC  FT DOMAIN 39 147 tRNA-BINDING.
CC CC  FT METAL 452 452 MAGNESIUM.
CC CC  FT METAL 458 458 MAGNESIUM (VIA CARBONYL OXYGEN).
CC CC  FT METAL 461 461 MAGNESIUM.
CC CC  FT METAL 462 462 MAGNESIUM.
CC CC  FT CONFLICT 369 370 QV -> AG (IN REF. 1).
CC CC  FT CONFLICT 372 376 AQRR -> RPEAG (IN REF. 1).
CC CC  FT CONFLICT 378 382 SLIQA -> KPPPG (IN REF. 1).
CC CC  FT STRAND 2 4
CC CC  FT HELIX 5 8

```



```

AC P4:028;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-PEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Putative stage IV sporulation protein.
GN SPOIV.
OS Bacillus megaterium.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1404;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 313;
RX MEDLINE=94288995; PubMed=7764969;
RA Meinhardt F., Busskamp M., Wittchen K.D.;
RT "Cloning and sequencing of the leu C and npr M genes and a putative
RT spo IV gene from Bacillus megaterium DSM319.";
RL Appl. Microbiol. Biotechnol. 41:344-351(1994).
CC -!- FUNCTION: MAY HAVE A ROLE IN COMPLETING STAGE IV OF SPOULATION.
CC -!- SIMILARITY: TO B.SUBTILIS YQFD.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X75069; CNA52963.1; -
DR PIR; I40228; I40228.
KW Sporulation.
SQ SEQUENCE 247 AA; 28319 MW; 05D742B5264BE066 CRC64;
Query Match 52.3%; Score 34; DB 1; Length 247;
Best Local Similarity 54.5%; Pred. No. 23; Mismatches 0; Gaps 0;
Matches 6; Conservative 3; Indels 2;
QY 1 APLGVQRAQFQ 11
DB 130 AKIGVKRGEFQ 140

RESULT 15
YBFF EC057
ID YBFF EC057 STANDARD; PRT; 375 AA.
AC Q8X605;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical RNA methyltransferase ybJF (RC 2.1.1.-).
GN YBFF OR Z1086 OR ECS0939.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,

```

```

RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -!- FUNCTION: Could be a 23S rRNA (Uracil-5)-methyltransferase.
CC -!- SIMILARITY: BELONGS TO THE RNA M5U METHYLTRANSFERASE FAMILY. YBFF
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AE905266; AAG55235.1; -
DR EMBL; AF002553; BAB34362.1; -
DR PIR; C90746; G85596.
DR HAMAP; MF 01012; -; 1.
DR InterPro; IPR000051; SAM_bind.
DR InterPro; IPR001566; TrmA.
DR PROSITE; PS01230; TRMA_1; 1.
DR PROSITE; PS01231; TRMA_2; 1.
DR KW Hypothetical protein; Transferase; Methyltransferase; rRNA processing;
KW Complete proteome.
FT ACT_SITE 334 334 BY SIMILARITY.
SQ SEQUENCE 375 AA; 41960 MW; 53C8E793B15BE05B CRC64;
Query Match 52.3%; Score 34; DB 1; Length 375;
Best Local Similarity 58.3%; Pred. No. 36;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 1 APLGVQRAQFOS 12
DB 276 AEGLTRLOFQA 287

Search completed: January 20, 2004, 16:54:21
Job time : 7.84733 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:42:36 ; Search time 37.5115 Seconds
(without alignments)
89.431 Million cell updates/sec

Title: US-10-089-549-18

Perfect score: 65

Sequence: 1 APLGVQRAQFQSG 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first: 45 summaries

Database : SPTREMBL_23.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	60.0	307	12 Q86756	Q86756 mucosal dis
2	39	60.0	330	12 Q02634	Q02634 mucosal dis
3	39	60.0	346	3 Q9C2Y1	Q9C2Y1 botrytis ci
4	39	60.0	354	12 Q9IN46	Q9IN46 bovine vira
5	39	60.0	963	12 Q9IU70	Q9IU70 bovine vira
6	39	60.0	1130	12 Q71955	Q71955 bovine vira
7	39	60.0	3898	12 Q9IWA6	Q9IWA6 mucosal dis
8	39	60.0	3898	12 Q11994	Q11994 mucosal dis
9	39	60.0	3907	12 Q96662	Q96662 mucosal dis
10	39	60.0	3975	12 Q65815	Q65815 mucosal dis
11	39	60.0	3988	12 Q9IWA7	Q9IWA7 mucosal dis
12	39	60.0	4040	12 Q9IFH8	Q9IFH8 mucosal dis
13	39	60.0	4983	12 Q11993	Q11993 mucosal dis
14	38	58.5	706	16 Q8PR40	Q8PR40 xanthomonas
15	38	58.5	787	5 Q9UA44	Q9UA44 aplysia cal
16	37	56.9	136	16 Q8FZB8	Q8FZB8 brucella su

17	37	56.9	184	16 Q8XZE6	Q8XZE6 ralstonia s
18	37	56.9	276	16 Q9HW28	Q9HW28 pseudomonas
19	37	56.9	1080	12 Q8QP47	Q8QP47 mucosal dis
20	36.5	56.2	1144	2 Q45745	Q45745 bacillus th
21	36	55.4	132	5 Q02613	Q02613 crithidia f
22	36	55.4	229	11 Q8C466	Q8C466 mus musculu
23	36	55.4	242	12 Q96737	Q96737 avian adeno
24	36	55.4	244	12 Q96729	Q96729 fowl adeno
25	36	55.4	246	12 Q85276	Q85276 avian adeno
26	36	55.4	252	16 Q8Y0P9	Q8Y0P9 ralstonia s
27	36	55.4	381	12 Q9YMM5	Q9YMM5 lymantria d
28	36	55.4	399	16 Q9KVD1	Q9KVD1 vibrio chol
29	36	55.4	653	10 Q8W5H2	Q8W5H2 oryza sativ
30	36	55.4	664	16 Q8PP79	Q8PP79 xanthomonas
31	36	55.4	772	5 Q9VF10	Q9VF10 drosophila
32	36	55.4	977	5 Q9VS83	Q9VS83 drosophila
33	36	55.4	979	5 Q8MQJ3	Q8MQJ3 drosophila
34	36	55.4	1137	2 Q9AGH9	Q9AGH9 pseudomonas
35	36	55.4	1332	5 Q9VS82	Q9VS82 drosophila
36	36	55.4	2051	5 Q44328	Q44328 hirudo medi
37	35	53.8	45	16 Q8FW01	Q8FW01 brucella su
38	35	53.8	48	12 Q9WLK4	Q9WLK4 hepatitis e
39	35	53.8	149	12 Q56047	Q56047 hepatitis e
40	35	53.8	149	12 Q9W9E8	Q9W9E8 hepatitis e
41	35	53.8	150	12 Q57042	Q57042 hepatitis e
42	35	53.8	184	12 Q9YPB0	Q9YPB0 hepatitis e
43	35	53.8	188	2 Q8RR68	Q8RR68 actinobacil
44	35	53.8	237	16 Q9RW82	Q9RW82 deinococcus
45	35	53.8	245	12 Q10644	Q10644 mucosal dis

ALIGNMENTS

RESULT 1

Q86756 PRELIMINARY; PRT; 307 AA.
AC Q86756; (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE GP48 (Fragment).
GN GP48.
OS Mucosal disease virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Pestivirus.
OX NCBI_TaxID=110399;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94262276; PubMed=82031115;
RA Silva-Krott I.U., Kennedy M.A., Potgieter L.N.;
RT *Cloning, sequencing, and in vitro expression of glycoprotein gp48 of
RL a noncytopathogenic strain of bovine viral diarrhoea virus.*;
RL Vet. Microbiol. 39:1-14 (1994).
DR EMBL: S71487; AAB30927.1;
DR InterPro: IPR001568; RNase T2.
DR PROSITE; PS00531; RNASE_T2_2, 1.
FT NON TER 307 307
SQ SEQUENCE 307 AA; 34539 MW; 43126ED9945E4F8 CRC64;

Query Match 60.0%; Score 39; DB 12; Length 307;
Best Local Similarity 70.0%; Pred. No. 24;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 GVQRAQFQSG 13
|:|:|:|:|
Db 94 GIORAMFQSG 103

RESULT 2

Q02634 PRELIMINARY; PRT; 330 AA.
ID Q02634
AC Q02634;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein (Fragment).
 OS Mucosal disease virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Pestivirus.
 OX NCBI_TaxID=11099;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Pe515CP;
 RX MEDLINE=9303130; PubMed=132926;
 RA Meyers G.; Tautz N., Stark R., Brownlie J., Dubovi E.J., Collett M.S.,
 RA Thiel H.-J.;
 RT "Rearrangement of viral sequences in cytopathogenic pestiviruses.";
 RL Virology 191:368-386(1992).
 DR EMBL; M96638; AAA42858.1; -;
 KW Nonstructural protein.
 FT NON TER 330 330
 SQ SEQUENCE 330 AA; 37127 MW; D958F23F6F771398 CRC64;
 Query Match 60.0%; Score 39; DB 12; Length 330;
 Best Local Similarity 70.0%; Pred. No. 26;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 4 GVQRAQFQSG 13
 Db |::|||
 285 GIQAMFQFG 294
 RESULT 3
 Q9C2Y1 PRELIMINARY; PRT; 346 AA.
 AC Q9C2Y1;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
 DE Pectin methyl esterase (EC 3.1.1.11).
 GN BCPME1.
 OS Botrytis cinerea (Botryotinia fuckeliana).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;
 OC Helotiales; Sclerotiniaceae; Botryotinia.
 OX NCBI_TaxID=40559;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=T4;
 RA Valette O., Levis C., Reignault P., Boccara M.;
 RT "Pectin methyl esterase contributes to Botrytis cinerea pathogenicity
 RT : interruption of bcpme gene reduces its attack on several plant
 RT hosts.";
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ309701; CAC29255.1; -;
 DR InterPro; IPR003070; Pectinesterase.
 DR Pfam; PF01095; Pectinesterase; 1.
 DR PROSITE; PS00503; PECTINESTERASE_2; 1.
 KW Hydrolase.
 SQ SEQUENCE 346 AA; 37236 MW; A1D81FC70C930F67 CRC64;
 Query Match 60.0%; Score 39; DB 3; Length 346;
 Best Local Similarity 80.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 4 GVQRAQFQSG 13
 Db |::|||
 311 GVQRAQFQSG 320
 RESULT 4
 Q91N46 PRELIMINARY; PRT; 354 AA.
 ID Q91N46
 AC Q91N46;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Polyprotein (Fragment).
 OS Bovine viral diarrhea virus-1.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Pestivirus.
 OX NCBI_TaxID=121864;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCP 1741;
 RX MEDLINE=21306326; PubMed=11413291;
 RA Becher P., Orlich M., Thiel H.J.;
 RT "RNA Recombination between Persisting Pestivirus and a Vaccine Strain:
 RT Generation of Cytopathogenic Virus and Induction of Lethal Disease.";
 RL J. Virol. 75:6256-6264(2001).
 DR EMBL; AF321453; AAK69178.1; -;
 DR InterPro; IPR001568; RNase T2.
 DR PROSITE; PS00531; RNASE_T2_2; 1.
 FT NON TER 354 354
 SQ SEQUENCE 354 AA; 40185 MW; CC28AF74768557DC CRC64;
 Query Match 60.0%; Score 39; DB 12; Length 354;
 Best Local Similarity 70.0%; Pred. No. 28;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 4 GVQRAQFQSG 13
 Db |::|||
 285 GIQAMFQFG 294
 RESULT 5
 Q91U70 PRELIMINARY; PRT; 963 AA.
 ID Q91U70
 AC Q91U70;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Polyprotein (Fragment).
 OS Bovine viral diarrhea virus strain Braidwood.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Pestivirus.
 OX NCBI_TaxID=159477;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Braidwood;
 RA Frost M.J., Papa R.A., Mackintosh S.G., Gu X., Shannon A.D.;
 RT "Australian pestivirus sequences.";
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF255049; AAK56854.1; -;
 DR InterPro; IPR001568; RNase T2.
 DR PROSITE; PS00531; RNASE_T2_2; 1.
 FT NON TER 963 963
 SQ SEQUENCE 963 AA; 108642 MW; E5ED2906EB62A7FF CRC64;
 Query Match 60.0%; Score 39; DB 12; Length 963;
 Best Local Similarity 70.0%; Pred. No. 82;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 4 GVQRAQFQSG 13
 Db |::|||
 115 GIQAMFQFG 124
 RESULT 6
 O71955 PRELIMINARY; PRT; 1130 AA.
 ID O71955
 AC O71955;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Polyprotein (Fragment).
 OS Bovine viral diarrhea virus strain Trangle Y546.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Pestivirus.
OX NCBI_TaxID=158474;
RN [1]
R2 SEQUENCE FROM N.A.
RC STRAIN=Trangie Y546;
RA Gu X., Frost M.J., Mackintosh S.G., Papa R.A., Shannon A.D.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF049222; AAC06279.2; -.
DR InterPro; IPR001356; Homeobox.
DR PROSITE; PS00027; HOMEBOX.1; 1.
DR PROSITE; PS00531; RNASE_T2.2; 1.
FT CHAIN 271 >97 E0 STRUCTURAL PROTEIN.
FT NON_TER 1130
SQ SEQUENCE 1130 AA; 127551 MW; FC227D159B1117F8 CRC64;

Query Match 60.0%; Score 39; DB 12; Length 1130;
Best Local Similarity 70.0%; Pred. No. 97;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 GVQRAQFQSG 13
Db 285 GIQRAMFQSG 294

RESULT 7
ID Q96662 PRELIMINARY; PRT; 3898 AA.
AC Q96662;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Mucosal disease virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Pestivirus.
OX NCBI_TaxID=11099;
RN [1]
R2 SEQUENCE FROM N.A.
RC STRAIN=type 1;
RA MEDLINE=20473135; PubMed=11018279;
RA Vassiliev V.B., Denis R.O.;
RT "Bovine viral diarrhoea virus induced apoptosis correlates with
RT increased intracellular viral RNA accumulation.";
RL Virus Res. 69:95-107(2000).
DR EMBL; AJ333739; CAB91847.1; -.
DR HSP; P27958; 1A1V.
DR MEROPS; S31.001; -.
DR InterPro; IPR000280; CDvir_endptseP80.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002166; HCV RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR001005; Myb DNA binding.
DR InterPro; IPR001568; RNase T2.
DR InterPro; IPR007095; RNA pol DS PS.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral RdRP; 1.
DR PRINTS; PR00729; CDVENDOPTASE.
DR SMART; SMO0487; DEXDC; 1.
DR PROSITE; PS00037; MYB_1; 1.
DR PROSITE; PS05057; RORP POSITIVE; 1.
DR PROSITE; PS050521; RORP VIRAL; 1.
DR PROSITE; PS00531; RNASE T2.2; 1.
DR ATP-binding; Helicase; Hydrolase; Nonstructural protein; Polyprotein;
DR RNA-directed RNA polymerase; Transferase.
SQ SEQUENCE 3898 AA; 438750 MW; 7F3995B13A13965 CRC64;

Query Match 60.0%; Score 39; DB 12; Length 3898;
Best Local Similarity 70.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 GVQRAQFQSG 13
Db 285 GIQRAMFQSG 294

RESULT 8
ID O11994 PRELIMINARY; PRT; 3898 AA.
AC O11994;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Pestivirus type 1 noncytopathic genomic RNA, complete genome (Genome
DE polyprotein).
OS Mucosal disease virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Pestivirus.
OX NCBI_TaxID=11099;
RN [1]
R2 SEQUENCE FROM N.A.
RC STRAIN=ILLNC;
RA Roath P.D., Berry E.S.;
RT "Nucleotide Sequence of ILLNC-Bovine Viral Diarrhea Virus.";
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U86600; A858572.1; -.
DR HSP; P27958; 1A1V.
DR InterPro; IPR000515; BPD transp.
DR InterPro; IPR000280; CDvir_endptseP80.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002166; HCV RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR001005; Myb DNA binding.
DR InterPro; IPR001568; RNase T2.
DR InterPro; IPR007095; RNA pol DS PS.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral RdRP; 1.
DR PRINTS; PR00729; CDVENDOPTASE.
DR SMART; SMO0487; DEXDC; 1.
DR SMART; SMO0490; HELIC; 1.
DR PROSITE; PS00402; BPD TRANSP INN_MEMBER; 1.
DR PROSITE; PS00037; MYB_1; 1.
DR PROSITE; PS05057; RORP POSITIVE; 1.
DR PROSITE; PS050521; RORP VIRAL; 1.
DR PROSITE; PS00531; RNASE T2.2; 1.
DR ATP-binding; Helicase; Hydrolase; Nonstructural protein; Polyprotein;
DR RNA-directed RNA polymerase; Transferase.
SQ SEQUENCE 3898 AA; 438538 MW; E76536F84P56C579 CRC64;

Query Match 60.0%; Score 39; DB 12; Length 3898;
Best Local Similarity 70.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 GVQRAQFQSG 13
Db 285 GIQRAMFQSG 294

RESULT 9
ID Q96662 PRELIMINARY; PRT; 3907 AA.
AC Q96662;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Mucosal disease virus, and
OS Bovine viral diarrhoea virus-1.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Pestivirus.
OX NCBI_TaxID=11099, 121864;
RN [1]
R2 SEQUENCE FROM N.A.

```

RC SPECIES=Mucosal disease virus; STRAIN=CP7;
RX MEDLINE=9712601; PubMed=8970985;
RA Meyers G., Tautz N., Becher P., Thiel H., Kuemmerer B.M.;
RT "Recovery of cytopathogenic and noncytopathogenic bovine viral
RL diarrhoea viruses from cDNA constructs.";
RN J. Virol. 70:8636-8613 (1996).
RP SEQUENCE FROM N.A.
RC SPECIES=Bovine viral diarrhoea virus-1; STRAIN=CP7-5A;
RX MEDLINE=20392142; PubMed=10933696;
RA Becher P., Orlich M., Thiel H.-J.;
RT "Mutations in the 5' nontranslated region of bovine viral diarrhoea
RL virus result in altered growth characteristics.";
RN J. Virol. 74:7884-7894 (2000).
RP SEQUENCE FROM N.A.
RC SPECIES=Bovine viral diarrhoea virus-1; STRAIN=CP7-5A;
RA Becher P., Orlich M.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; U63479; AAC55984.1; -.
DR EMBL; AF220247; AAG00378.1; -.
DR HSSP; P27958; 1A1V.
DR InterPro; IPR000280; CDvir_endptsep80.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002166; HCV RdRP.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR001005; Myb DNA Binding.
DR InterPro; IPR001568; RNase T2.
DR InterPro; IPR007094; RNA_pol_DS_PS.
DR InterPro; IPR007095; RNA_pol_PSVir.
DR PRINTS; PR00729; CDVENDOPTASE.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
DR PROSITE; PS00037; MYB_1; 1.
DR PROSITE; PS50507; RDRP_POSITIVE; 1.
DR PROSITE; PS50521; RDRP_VIRAL; 1.
DR PROSITE; PS00531; RNase_T2_2; 1.
DR PROSITE; PS00299; UBIQUITIN_1; 1.
DR PROSITE; PS50053; UBIQUITIN_2; 1.
DR ATP-binding; Helicase; Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase.
SQ SEQUENCE 3975 AA; 439101 MW; 9E4B019FF8042410 CRC64;

Query Match 60.0%; Score 39; DB 12; Length 3907;
Best Local Similarity 70.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 GVQRAQFQSG 13
DB 285 GIQRAMFQSG 294

RESULT 10
Q5815 PRELIMINARY; PRT; 3975 AA.
AC Q65815;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Mucosal disease virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Pestivirus.
OX NCBI_TaxID=11099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=type 1;
RX MEDLINE=20473135; PubMed=11018279;
RA Vassiliev V.B., Denis R.O.;
RT "Bovine viral diarrhoea virus induced apoptosis correlates with
RT increased intracellular viral RNA accumulation.";
RL Virus Res. 69:95-107 (2000).
DR EMBL; AJ133738; CAB91846.1; -.
DR HSSP; P27958; 1A1V.
DR MEROPS; S31.001; .
DR InterPro; IPR000280; CDvir_endptsep80.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002166; HCV RdRP.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR001005; Myb DNA Binding.
DR InterPro; IPR001568; RNase_T2.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR PRINTS; PR00729; CDVENDOPTASE.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
DR PROSITE; PS00037; MYB_1; 1.
DR PROSITE; PS50507; RDRP_POSITIVE; 1.
DR PROSITE; PS50521; RDRP_VIRAL; 1.
DR PROSITE; PS00531; RNase_T2_2; 1.
DR PROSITE; PS00299; UBIQUITIN_1; 1.
DR PROSITE; PS50053; UBIQUITIN_2; 1.
DR ATP-binding; Helicase; Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase.
SQ SEQUENCE 3975 AA; 446078 MW; C7D83ADF3F0BD025 CRC64;

Query Match 60.0%; Score 39; DB 12; Length 3975;
Best Local Similarity 70.0%; Pred. No. 3.8e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 GVQRAQFQSG 13
DB 285 GIQRAMFQSG 294

RESULT 11
Q91WA7 PRELIMINARY; PRT; 3988 AA.
AC Q91WA7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Mucosal disease virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Pestivirus.
OX NCBI_TaxID=11099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=type 1;
RX MEDLINE=20473135; PubMed=11018279;
RA Vassiliev V.B., Denis R.O.;
RT "Bovine viral diarrhoea virus induced apoptosis correlates with
RT increased intracellular viral RNA accumulation.";
RL Virus Res. 69:95-107 (2000).
DR EMBL; AJ133738; CAB91846.1; -.
DR HSSP; P27958; 1A1V.
DR MEROPS; S31.001; .
DR InterPro; IPR000280; CDvir_endptsep80.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002166; HCV RdRP.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR001005; Myb DNA Binding.
DR InterPro; IPR001568; RNase_T2.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR PRINTS; PR00729; CDVENDOPTASE.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
DR PROSITE; PS00037; MYB_1; 1.
DR PROSITE; PS50507; RDRP_POSITIVE; 1.
DR PROSITE; PS50521; RDRP_VIRAL; 1.
DR PROSITE; PS00531; RNase_T2_2; 1.
DR PROSITE; PS00299; UBIQUITIN_1; 1.
DR PROSITE; PS50053; UBIQUITIN_2; 1.
DR ATP-binding; Helicase; Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase.
SQ SEQUENCE 3975 AA; 446078 MW; C7D83ADF3F0BD025 CRC64;

```

DR SMART; SM00487; DEXDC; 1.
 DR SMART; SM00490; HELIC; 1.
 DR PROSITE; PS00037; MYB 1; 1.
 DR PROSITE; PS00507; RDRP POSITIVE; 1.
 DR PROSITE; PS00521; RDRP VIRAL; 1.
 DR PROSITE; PS00531; RNASE T2; 1.
 KW ATP-binding; Helicase; Hydrolase; Nonstructural protein; Polyprotein;
 RNA-directed RNA polymerase; Transferase.
 SQ SEQUENCE 3988 AA; 449122 MM; 8DCCF24D64C04A68 CRC64;

Query Match 60.0%; Score 39; DB 12; Length 3988;
 Best Local Similarity 70.0%; Pred. No. 3.8e+02;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 GVQRAQFQSG 13
 DB 285 GIQRAWFQSG 294

RESULT 12
 Q9IFH8 PRELIMINARY; PRT; 4040 AA.
 ID Q9IFH8
 AC Q9IFH8
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Genome polyprotein.
 OS Mucosal disease virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Pestivirus.
 OX NCBI_TaxID=11099;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20323484; PubMed=10864644;
 RA Lai V.C., Zhong W., Skelton A., Ingravallo P., Vassilev V.,
 RA Denis R.O., Hong Z., Lau J.Y.;
 RT "Generation and characterization of a hepatitis C virus NS3 protease-
 RT dependent bovine viral diarrhea virus.";
 RL J. Virol. 74:6339-6347 (2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Lai V.C.H., Hong Z.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF268278; AAF82566.1; -;
 DR HSSP; P26663; IUXP.
 DR MEROPS; S31.001; -;
 DR InterPro; IPR000280; CDvir_endptsep80.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR004109; HCV NS3.
 DR InterPro; IPR002166; HCV RdRp.
 DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR001005; Myb DNA Binding.
 DR InterPro; IPR001568; RNase T2.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF00998; Viral RdRp; 1.
 DR PRINTS; PR00729; CDVENDOPTASE.
 DR SMART; SM00487; DEXDC; 1.
 DR SMART; SM00490; HELIC; 1.
 DR PROSITE; PS00037; MYB 1; 1.
 DR PROSITE; PS00507; RDRP POSITIVE; 1.
 DR PROSITE; PS00521; RDRP VIRAL; 1.
 DR PROSITE; PS00531; RNASE T2; 1.
 KW ATP-binding; Helicase; Hydrolase; Nonstructural protein; Polyprotein;
 RNA-directed RNA polymerase; Transferase.
 SQ SEQUENCE 4040 AA; 453073 MM; ADE87791D055B9DC CRC64;

Query Match 60.0%; Score 39; DB 12; Length 4040;
 Best Local Similarity 70.0%; Pred. No. 3.9e+02;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 GVQRAQFQSG 13
 DB 337 GIQRAWFQSG 346

RESULT 13
 O11993 PRELIMINARY; PRT; 4983 AA.
 ID O11993
 AC O11993
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Pestivirus type 1 cytopathic genomic RNA, complete genome (Genome
 DE polyprotein).
 OS Mucosal disease virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Pestivirus.
 OX NCBI_TaxID=11099;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ILLC;
 RA Roath P.D., Betty E.S.;
 RT "Nucleotide Sequence of ILLC-Bovine Viral Diarrhea Virus.";
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U86599; AAB58571.1; -;
 DR HSSP; P02248; IUBI.
 DR MEROPS; C53.001; -;
 DR InterPro; IPR000280; CDvir_endptsep80.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002166; HCV RdRp.
 DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR001005; Myb DNA Binding.
 DR InterPro; IPR001568; RNase T2.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR InterPro; IPR000626; Ubiquitin.
 DR Pfam; PF00271; helicase_C; 2.
 DR Pfam; PF00240; ubiquitin; 1.
 DR Pfam; PF00998; Viral RdRp; 1.
 DR PRINTS; PR00729; CDVENDOPTASE.
 DR PRINTS; PR00348; UBIQUITIN.
 DR SMART; SM00487; DEXDC; 2.
 DR SMART; SM00490; HELIC; 2.
 DR SMART; SM00213; UBQ; 1.
 DR PROSITE; PS00037; MYB 1; 2.
 DR PROSITE; PS00507; RDRP POSITIVE; 1.
 DR PROSITE; PS00521; RDRP VIRAL; 1.
 DR PROSITE; PS00531; RNASE T2; 1.
 DR PROSITE; PS00531; UBIQUITIN_2; 1.
 KW ATP-binding; Helicase; Hydrolase; Nonstructural protein; Polyprotein;
 RNA-directed RNA polymerase; Transferase.
 SQ SEQUENCE 4983 AA; 558890 MM; 871A30A0B7D70A27 CRC64;

Query Match 60.0%; Score 39; DB 12; Length 4983;
 Best Local Similarity 70.0%; Pred. No. 4.8e+02;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 GVQRAQFQSG 13
 DB 285 GIQRAWFQSG 294

RESULT 14
 Q8PR40 PRELIMINARY; PRT; 706 AA.
 ID Q8PR40
 AC Q8PR40
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Iron transporter.
 DE YNCD OR XAC0126.
 OS Xanthomonas axonopodis (pv. citri).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 PLGVQRAQFQ 11
Db 6 PLDIQRLQFQ 15

Search completed: January 20, 2004, 17:01:21
Job time : 38.5115 secs

OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13962 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Perro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.N.C., do Anaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Clapina L.P.,
RA Cicarelli R.M.B., Coutinho L.B., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinoza L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setuba C.C., Kitajima J.P.;
RT *Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities;
RL Nature 417:459-463(2002).
DR EMBL; AE011635; AM35018.1; --
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Rec; 1.
XW Complete proteome.
SQ SEQUENCE 706 AA; 76546 MW; DB92B347CD6386B9 CRC64;

Query Match 58.5%; Score 38; DB 16; Length 706;
Best Local Similarity 63.6%; Pred. No. 92;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 PLGVQRAQFQ 12
Db 249 PLGLTRAQFNA 259

RESULT 15
Q9UA44
ID Q9UA44 PRELIMINARY; PRT; 787 AA.
AC Q9UA44;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Neutral endopeptidase.
GN NEP.
OS Aplysia californica (California sea hare).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Orthogastropoda;
OC Apogastropoda; Heterobranchia; Euthyneura; Opisthobranchia; Anaspidea;
OC Aplysioidea; Aplysiidae; Aplysia.
OX NCBI_TaxID=6500;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99274733; PubMed=10341232;
RA Zapallia J.P., Wickham L., Bawab W., Yang X.F., Storozhuk M.V.,
RA Castellucci V.F., DesGroselliers L.;
RT "Cloning and characterization of Aplysia neutral endopeptidase, a
RT metallo-endopeptidase involved in the extracellular metabolism of
RT neuropeptides in Aplysia californica."
RL J. Neurosci. 19:4280-4292(1999).
DR EMBL; AF104361; RAD51382.1; --
DR HSSP; P08473; 1DMT.
DR InterPro; IPR000718; Peptidase M13.
DR InterPro; IPR006025; Zn MTpeptidse.
DR Pfam; PF01431; Peptidase M13; 1.
DR PRINTS; PR00786; NEPRILYSIN.
DR PROSITE; PS00142; ZINC PROTEASE; 1.
SQ SEQUENCE 787 AA; 83171 MW; 3D91D1485CECE437 CRC64;

Query Match 58.5%; Score 38; DB 5; Length 787;
Best Local Similarity 70.0%; Pred. No. 1e+02;